



Data Integrity Initiative

A Toolkit from the Insights Association
In Partnership with a Council of Industry Leaders





TABLE OF CONTENTS

Introduction	3
Resources	
Awareness	5
Glossary	6
Data Quality Measurement & Evaluation	7
Professional Standards and Certifications.....	8
Advocacy and Policy	9
Glossary of Terms.....	10
Guidelines for Assessing Quality.....	14
What's Next?	15

INTRODUCTION

The Marketing Research industry experienced 16.6% revenue growth in 2021 versus year ago, posting significant gains and recovery following the impact that COVID had on the industry over past years (Brereton, et al., 2022). Segments, such as Self Service Platforms, Panel Providers, Enterprise Management Feedback, Digital Data & Analytics and Social Media Listening Platforms, experienced double-digit growth and magnified accelerating trends toward platform and sample-based approaches. In fact, insights are increasingly performed by a brand's internal non-research resources, who opt for user-driven tools and "roughly right" speed to insights over expert researchers and traditional, slower and more accurate research approaches. As a result, ensuring both providers and users of sample and data services are well versed on data integrity is of the utmost importance to preserving sound decision-making and the reputation and value of the marketing research industry.

As the Market Research industry expanded so has online fraud. An independent study of 50 online sample providers conducted by the PEW Research Center found a high rate of respondents that they described as "bogus" because of just how unusable their data are (Kennedy et al., 2020). The biggest concern is that the presence of such respondents in samples can completely invalidate the results of otherwise well-designed studies (Litman et al., 2021).

The increase in fraudsters attempting to participate in online research surveys shouldn't be surprising. While profit-seekers have long been active, their prevalence has increased during the pandemic. The Federal Trade Commission reports that consumers lost more than \$5.8 billion to fraud in 2021, an increase of 70% versus year ago. (FTC Report, 2022).

Fraudsters have become sophisticated and determined. Whether they are people who are just trying to feed their families,

enterprising professionals who teach others how to take surveys, organized groups running a cottage industry, or thieves stealing identities, fraudsters are disruptive and put data integrity at risk. Combating survey fraud - prevent fraudsters from taking surveys in the first place, as well as in evaluating and cleaning the data - is costly in terms of time, resources, and financial investment. Challenges increase as the use of data becomes democratized. End-users of DIY platforms do not have access to Insights Professionals who are aware of potential data issues and can evaluate the impact of data quality on research outcomes. In fact, the majority of entry-level job postings mentioning marketing research skills were not for traditional marketing research agencies, but instead, for marketing brand-side generalists who would conduct research only part of the time (Reavey, Rosenbloom and Zahay, 2021).

A lot of work has been done to educate the Insights & Analytics industry on how to identify and minimize fraud and improve data quality. Several industry resources exist, but they are not centralized and easily accessible. In addition, these resources may not be readily available to non-traditional researchers who are now being asked to conduct marketing research studies using DIY research tools. And while many sample and data providers have rigorous processes for cleaning data and preventing sample fraud, these practices vary in their effectiveness and by provider, leaving even the most seasoned researchers confused and uncertain on how to properly assess data quality with their vendors.

Advocacy and policy also play an important role in preserving data quality and the long-term integrity and viability of the Insights industry. The Insights Association, and its predecessors, have been advocating, and winning, on public policy issues impacting the insights industry for many years. This important work evolves from,



and contributes to, the rapidly changing marketing research environment and ensures steps needed to combat fraud, preserve data quality and protect the industry are in place.

In order to address many of these data quality concerns and opportunities in our industry, the Insights Association, in partnership with industry experts and contributors, is pleased to release the **Data Integrity Initiative (DII) Toolkit**. The goals of the DII Toolkit are to centralize industry information to educate and empower both researchers and end-users to be more aware and better equipped to address data quality issues. There is no one simple reason, cause or solution for data quality issues. Instead, it is up to all of us to pledge to stay informed, engaged and committed to ensuring data delivered by our industry meets the highest quality standards. This pledge is why so many leaders from a wide range of supplier, sample, client and even competitive companies have united and joined efforts to create this DII Toolkit for the industry. We have all taken the DII Pledge to ensure data and insights are reliable for decision-making. We now ask all of you to join us and take the DII Pledge by...

Raising awareness of sample quality, integrity and fraud by compiling industry best practices and guidance to benefit:

**Data & Sample
Providers**

**Insights
Consumers**

End Users

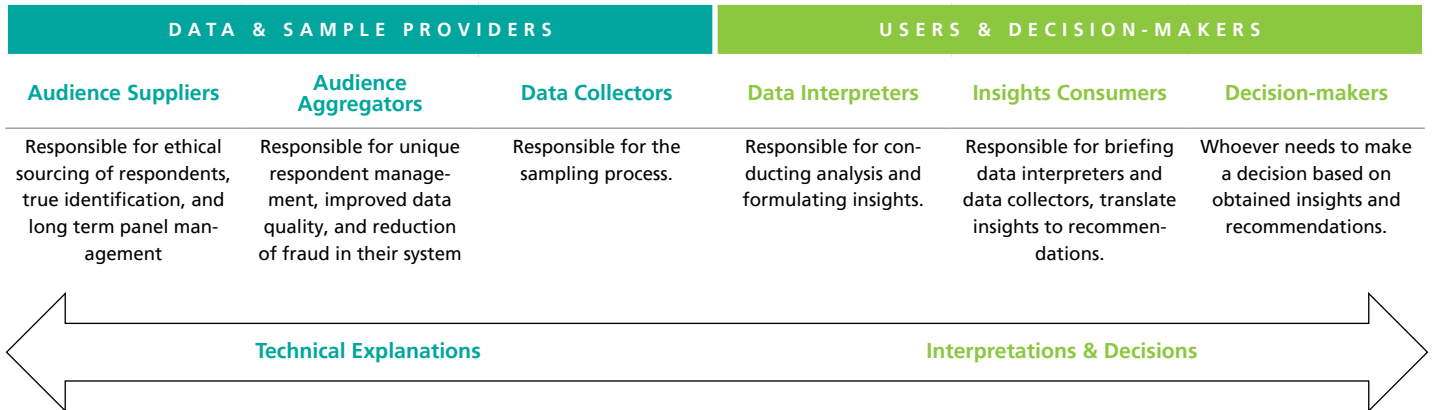
**Decision-
Makers**

And ensure reliable insights for confident design-making

The need to understand Data and Sample Integrity extends across all sectors of the Insights and Analytics industry, including *Data & Sample Providers to Users & Decision-Makers*. Building awareness of data integrity issues should be broad to include not only traditional *marketing research and sample provider* companies,

but also end-users across a wide variety of client-side industries and verticals. There is a need to elevate basic understanding and evaluation of data integrity issues for all end users, while at the same time, ensuring transparency and discipline among sample and data providers. To fully benefit the industry, sample and data integrity

definitions, quality measures and evaluation processes should continuously evolve and extend to as many research methodologies as needed. In addition, familiarity with professional standards and certifications is important for elevating overall knowledge of potential solutions for achieving data and sample integrity.



AWARENESS: Build Awareness of Data Quality & Integrity Issues and potential impact
DEFINE: Define Data Quality and Integrity Terms, Highlighting Implications on Results (Glossary)
MEASURES: Develop Quality Measures, Scorecards and Benchmarks to guide data evaluations

EVALUATE: Empower Users to Evaluate & Select Data Providers adhering to Quality Standards
PROFESSIONAL CERTIFICATIONS & STANDARDS: Explore professional certifications and standards and how they might benefit your work and organizations

ADVOCACY AND POLICY: Ensure policies and legislation are in place to preserve data quality and protect the industry in a rapidly changing marketing research and data environment

CALL TO ACTION

This is only the beginning, not the end, of working together to educate and empower all providers and users of marketing research on the actions needed to preserve data quality and integrity in the industry. As part of this journey, we invite industry leaders to unite and join the Insights Association DII Council to continue to evaluate, solve and share data integrity issues across the broad set of marketing research verticals, methods and user types. We also ask everyone involved in any phase of the research process to join us in taking the **DII Pledge to Raise Awareness** of data quality, integrity and fraud issues to *ensure Reliable Insights for Confident Decision-Making are available to all*.

We plan to continuously update this document with additional data quality resources and learnings, to ensure all

providers, suppliers and users of sample data are aware of the latest industry developments. Our work thus far applies to both quantitative and qualitative data, however we recognize that qualitative data integrity has unique challenges that deserve additional focus. As a result, the IA DII Council will examine the nuances of qualitative data quality as the next step in our Data Integrity efforts.

In addition, the Insights Association will collaborate with ESOMAR on a project to better understand the respondent experience of survey research.

PLEASE JOIN US and take the DII Pledge by downloading the DII Toolkit, adopting DII quality elements into your research practices, and training your teams to strive for quality and excellence in all research.

The IA Data Integrity Task Council

- Michael Brereton: Michigan State University
- Bonnie Breslauer: DISQO
- Melanie Courtright: Insights Association
- Nick Flores: Imperium
- Brad Franz: Burke
- Anouar El Haji, PhD.: Veylinx
- Kerry Hecht: Echo
- Dawn Hoskins Fite: MarketVision Research
- AJ Keirans: Paradigm Sample
- Vignesh Krishnan: Research Defender
- Leib Litman: CloudResearch
- Jamie Lussier: Travelers
- James Lytle: Big-Village
- Cindy Neumann: Insights Association
- Katrina Noelle: KNow Research
- Sylvia Perez: DMI Consulting
- Deb Ploskonka: Cambia Information Group
- Cheskie Rosenzweig: CloudResearch
- Juliana Wood: CIRQ



AWARENESS

Below are resources designed to build awareness of data and sample integrity issues in the industry

Insights Association Town Hall: Sample Fraud: Causes, Costs, Cures (CASE)

[CLICK HERE](#)

Insights Association Town Hall: Sample Fraud: Continuing the Discussion (CASE)

[CLICK HERE](#)

Insights Association Town Hall: Data Quality: Combatting New Threats (CASE)

[CLICK HERE](#)

Insights Association Town Hall: Brands Unite to Disrupt the Status Quo of Data Quality (CASE)

[CLICK HERE](#)

Insights Association Mid-Atlantic Symposium: Detecting and Eliminating Fraud in Survey Research - Join the Effort (CASE)

[CLICK HERE](#)

Insights Association Combined Meeting of Philadelphia Chapters and QRCA: Quality and Qualitative - Should Brands Be Concerned (CASE)

[CLICK HERE](#)

Assessing the Risks to Online Polls From Bogus Respondents (Pew Research)

[CLICK HERE](#)

Did people really drink bleach to prevent COVID-19? A tale of problematic respondents and a guide for measuring rare events in survey data (mexrxiv)

[CLICK HERE](#)

What's Next - Future Work

[CLICK HERE](#)



GLOSSARY

The IA DII Council has reviewed industry sources and developed a comprehensive Glossary of terms and definitions related to data integrity across seven key areas:

1. Respondent characteristics relating to data quality
2. Behavior and biases relating to problematic respondents
3. Tools and methods for detecting problematic respondents
4. Characteristics of panels and samples that can affect data quality
5. Issues relating to survey design that can affect data quality
6. Terms relating to the validation of fraud-detection solutions
7. General terms

[IA Data Integrity Initiative Glossary of Terms](#)

[CLICK HERE](#)



DATA QUALITY MEASUREMENT & EVALUATION

Below are Measurement and Evaluation approaches that can be used within a “Code of Conduct” for best practices across our industry. Adoption across broad stakeholder groups allows for continuous improvement across all data quality areas. Guides on selecting data suppliers and providers, designed specifically for end-users and decision-makers, are also shown below. In addition, the Insights Association’s DII Council has created the Checks of Integrity framework that can guide providers and users of research on key data integrity measures and evaluations needed across all phases of survey research.

[Insights Association: Code for Standards & Ethics](#)

[CLICK HERE](#)

[ESOMAR/GRBN Guideline for Online Sample Quality](#)

[CLICK HERE](#)

[Engage 2.0: 30 Tips to Improve the Research Participant User Experience](#)

[CLICK HERE](#)

[DII Guidelines for Assessing Quality](#)

[CLICK HERE](#)

[In-Survey Checks](#)

[CLICK HERE](#)

[Defeating Endemic Cheating - Cambia Information Group](#)

[CLICK HERE](#)

[CASE: Recommended Questions to Ask Suppliers](#)

[CLICK HERE](#)

[How Brands Are Uniting to Ensure Data Quality - LUCY](#)

[CLICK HERE](#)

[AAPOR: Evaluating Survey Quality in Today’s Complex Environment](#)

[CLICK HERE](#)

[ESOMAR 36 Answers](#)

[CLICK HERE](#)

[Quantitative Data Best Practices](#)

[CLICK HERE](#)

[In-Survey Quality Data Checks: One Tool in the Quality Toolbox](#)

[CLICK HERE](#)

CHECKS OF INTEGRITY

BEHAVIORAL VALIDATIONS

PRE-SURVEY

- Geo-Location Tracking
- Digital Device Fingerprinting
- IP-Deduplication
- Bots
- Bot Script
- Click-Farm

IN-SURVEY

- Form Filler Check
- Instructional Manipulation Check
- Low Incidence Check
- Speeder/Racer
- Straightlining
- Red Herring/Trap Question
- Contradictory Answers
- Open-End Response Validation
- Survey

POST-SURVEY

- Data Discrepancy Checks
- Open-End Response Review
- Benchmark Comparison
- MaxDiff Question Check



PROFESSIONAL STANDARDS AND CERTIFICATIONS

Organizations in the market insights space often have well-defined processes. The path to certification adds a level of increased compliance to the quality research management system and lets your customers know that adherence to the global market research and insights standard is a priority that you practice every day, for every client, for every project.

Users of data and information also benefit from understanding the benefits of working with suppliers who are certified and follow proven industry standards. In addition, pursuing their own certifications and adoption of industry standards can elevate the quality of data and information shared with their organizations and ensure decision-making is based on the best and most accurate information possible.

[Insights Association: ISO Certification webinars and other information on ISO Certifications and the certification process](#)

[CLICK HERE](#)

[Insights Association Webinar: ISO Certifications and Processes for Achieving Certification](#)

[CLICK HERE](#)

RESOURCES AND DESCRIPTION

THE STANDARD

ISO 20252:2019 is the global market research and insights standard recognized around the world. The expanded and modernized version provides operational frameworks that help ensure consistent levels of quality across offices, project teams and suppliers. The ISO framework sets parameters to ensure consistent supplier process and quality control measures – a huge plus for companies that require outside data collection and programming.

Insights and data analytics providers that align their research process to ISO 20252 will experience fewer errors and unnecessary duplication, improved quality levels, enhanced repeatability and productivity, and easier progress measurement. For human resources operations, the onboarding and training of new employees becomes more effective with this quality framework as a guide.

CERTIFICATION INSTITUTE FOR RESEARCH QUALITY (CIRQ)

Securing certification through CIRQ, the certification body of the Insights Association, shows your organization's commitment to the most up-to-date research practices and technologies through ongoing compliance and annual external audits, your commitment to excellence and can give your company a keen edge in the competitive research landscape. CIRQ was established in compliance with all ISO requirements for certification bodies that provide auditing and certification services and is fully accredited by ANSI's National Accreditation Board.



ADVOCACY AND POLICY

Below are advocacy and policy resources on a wide variety of topics that impact data quality and protect the marketing research industry overall.

Advocacy Successes: IA's efforts have resulted in successful legislative wins that benefit the Insights Industry in the following areas:

[CLICK HERE](#)

[Consumer Privacy and Data Security](#)

[Research Participants = Independent Contractors](#)

[Healthcare in Marketing Research](#)

[Data Collection](#)

Legal Affairs: The Insights Association provides legal/compliance information on issues related to the Insights industry, often by state and industry sector:

[CLICK HERE](#)

[Consumer Privacy and Data Security](#)

[Respondent Incentives](#)

[Telephone, Fax and Email](#)

[Pharmaceutical, Medical Devices, and Healthcare](#)

[International](#)

[Business, Accounting and Labor Concerns](#)

[Model Contract Forms and Language](#)

Regulations & Standards:

[GDPR](#)

[CLICK HERE](#)

[CCPA](#)

[CLICK HERE](#)



GLOSSARY OF TERMS

RESPONDENT CHARACTERISTICS RELATING TO DATA QUALITY

FRAUD: Intentional misrepresentation of identity in a sample, deception or malicious behavior intended only for financial gain with no real insights provided.

FRAUDULENT RESPONDENT: An umbrella term denoting a respondent who intentionally circumvents the survey process for monetary gain. Most commonly, survey fraud occurs when respondents attempt to collect survey rewards while a) not being qualified for the survey and/or b) making efforts to collect rewards while bypassing as much of the survey as possible. This includes: 1) not responding to questions honestly, 2) taking surveys one is not qualified for, 3) falsely posing as belonging to a particular demographic group, 4) accessing a survey from a country not being targeted or in a language the survey is not being offered in, 5) intentionally taking the survey more than 1x from the same or multiple different accounts, 6) the use of automation to generate closed-ended and open-ended survey responses.

PROBLEMATIC RESPONDENT: An umbrella term which includes inattentive, mischievous, and other forms of bad quality respondents, including those who use certain types of automation such as translation apps. Not all forms of problematic respondents should be categorized as fraud. For example, inattention may be considered as problematic respondent behavior rather than fraud.

VALID RESPONDENT: A respondent who is attentive, engaged, honest, and meets survey participation requirements.

INATTENTIVE RESPONDENT: A respondent who is disengaged and responds to questions without fully processing or understanding their content.

MISCHIEVOUS RESPONDENT: Respondents who provide information that is intentionally false or misleading. As opposed to fraud, mischievousness is less likely to be motivated by financial gain. For example, 400,000 people reported their religion as Jedi in the 2001 British Census.

BAD ACTORS: Respondents intentionally misrepresenting who they are or what they do in order to qualify for a survey.

PROFESSIONAL SURVEY TAKER (PST): An individual who participates in surveys at a higher rate than most

Good intent: could simply be an (over)active participant that provides valid and quality responses

Bad intent: one simply participating in surveys at a high rate but with no regard for validity or quality of responses

BEHAVIOR AND BIASES RELATING TO PROBLEMATIC RESPONDENTS

ACQUIESCENCE BIAS (AKA YEA-SAYING): Respondents who tend to disproportionately select a Yes response over other available response options.

RANDOM RESPONSE PROFILE: When a respondent randomly selects from among a question's available response options. Note that this is different from straightlining or yea-saying, which are not random, and are more easily detectable. Form-fillers at times employ a random or semi-random response strategy.

POSITIONAL BIAS: When a respondent preferentially selects a response based on its position. Most commonly, problematic respondents select the first (or top) available response option.

FORM FILLER: An app, typically a browser plugin, which automatically fills out survey questions.

STRAIGHTLINING: Providing the same answer to the majority of grid questions. The threshold for what is considered straightlining is often set at 80%. However, the specific threshold can depend on the nature of the study.

SPEEDING/RACING: Extremely fast survey completion times. Thresholds for what is considered “very fast” can either be set relationally (e.g. fastest 10% of the sample; completion times below three standard deviations) or based on specific minimum completion times (e.g. faster than 1 minute in a study that takes ten minutes on average). Many studies show that some valid respondents are capable of very fast completion times, and care must be taken to avoid false positives. Thus, speeding is often used as a flag, rather than as a sole rejection criterion.

NON-NAIVETE: Response bias that results from familiarity with the survey intentions or content. This is most commonly observed among highly active respondents who participate in many studies.

BOTS: Computer programs designed to mimic human activity to participate in online surveys for purposes the incentive or reward.

BOT SCRIPT: Individual accessing a survey where computer software is then used to auto-fill the survey.

CLICK FARM: Significant activity from group of respondents generally for malicious purposes to participate in surveys and earn rewards.

TOOLS AND METHODS FOR DETECTING PROBLEMATIC RESPONDENTS

VERIFICATION: The process to verify the identity or background of a respondent through personal info or other qualifying questions.

ONGOING PANEL VALIDATION: Where you monitor panel targeting/profiling to make sure it's consistent with qualifying or termination metrics. Would require information passed back to know more on where respondents are qualifying or terming in a survey.

DUPLICATES - The same respondent completed the same survey more than one time, as identified by checks on IP addresses and browser fingerprints, providing the same responses throughout a survey, or providing the same PII when that information is collected. Intent may vary: is this someone purposely trying to take the survey multiple times or simply on multiple panels where they receive the same survey.

IP-DEDUPLICATION: Technology that detects respondents who are taking the survey from the same IP address.

DIGITAL DEVICE FINGERPRINTING: Technology that detects respondents who are taking the survey from the same device.

OPEN-ENDED RESPONSE VALIDATION: The process of examining open-ended responses to determine the quality of the participant

- **Copy/Paste:** the act of using previously scripted words or phrases from an outside source for in-survey answers.
- **Gibberish:** open end responses that provide no value or substance. Can include random key inputs or answers not tied to the question.
- **Non engaged:** brief or short answers that lack substance or value to the desired questions.

GEO-LOCATION TRACKING: Identifying the physical location of participants (usually via IP address) to ensure participants are in the geographic locale they claim to be in.

BEHAVIORAL VALIDATION: The process of identifying problematic respondents through an examination of their behavior, which can include responses to specific survey questions, response patterns across multiple questions, mouse movements and other behaviorometric techniques. Behavioral validation can be applied both pre-survey and in-survey.

INSTRUCTIONAL MANIPULATION CHECK (IMC) OR RED HERRING/TRAP QUESTION: A question in a survey designed to check whether respondents are paying attention. A common example instructs respondents to select a “strongly disagree” response option.

LOW INCIDENCE CHECK: A question designed to include a few low incidence/unlikely options that check for respondent's attention.

CONTRADICTORY ANSWERS OR DATA DISCREPANCY: Answers covering multiple questions that don't agree. Questions and answers should be very similar so as to not change the interpretation for the respondent.

PRE-SURVEY QUALITY VALIDATION: The process of identifying and removing low quality respondents before they enter a survey. This is one type of Behavioral Validation.

MAXDIFF QUESTIONS: Results from a MaxDiff (Maximum Difference or Best/Worst Scaling) exercise can also be used to identify patterns of fraudulent responses.

CHARACTERISTICS OF PANELS AND SAMPLES THAT CAN AFFECT DATA QUALITY

SUPPLIER: A company that provides researchers with access to participants willing to take part in surveys.

PANEL SAMPLE: Respondents who have registered with a particular site (panel) and indicated their interest in participating in surveys.

OPT-IN: To “opt-in” to an online panel is to sign up for the panel as a participant. This involves providing an email address and potentially other information.

DOUBLE OPT-IN: After signing up for a panel (opt-in) a respondent will be sent a confirmation email. Double opt-in involves confirming an email address by responding to the confirmation email. The double opt-in process does not by itself provide sufficient protection against fraud.

RIVER SAMPLE: Respondents who participate in surveys via banners, video games and other ads. There is typically no opt-in process. River samples improve reach, but more care is required to ensure data quality.

AGGREGATOR: A company that provides access to participants by gathering multiple panel sources and making them all accessible via a single interface.

ROUTER: Technology that redirects respondents to specific surveys.

B2B: “Business to Business” surveys that are targeted toward business professionals. Examples of commonly targeted B2B groups include IT decision makers (ITDM), human resources decision makers (HRDM), and Healthcare providers. In addition to the types of fraud that threaten any market research study, B2B surveys are particularly vulnerable to false claims of group membership that can severely compromise the validity of conclusions. Fraud rates in B2B studies are often much higher than in B2C studies, in part because of the higher monetary rewards associated with such studies.

B2C: “Business to Consumer” surveys that are targeted toward the general consumer.

ISSUES RELATING TO SURVEY DESIGN THAT CAN AFFECT DATA QUALITY

RESPONDENT EXPERIENCE: How a respondent feels when participating in a survey. A respondent can experience excitement, boredom, interest, frustration, and a wide range of other reactions that can affect their level of engagement, attention, and focus, influencing the quality of their responses.

SURVEY DESIGN: Elements of survey methodology that affect how and in what manner stimuli are presented to the respondent. Various aspects of survey design can affect respondent experience, including length of interview, the use of matrix questions, the use of a progress bar, clear instructions, and numerous other design considerations.

LENGTH OF INTERVIEW (LOI): Median survey completion time.

MATRIX (AKA GRID) QUESTIONS: Closed-ended survey questions with a characteristic gridlike column structure in which columns typically correspond to response-options.

TERMS RELATING TO THE VALIDATION OF FRAUD-DETECTION SOLUTIONS

VALIDATION OF FRAUD-DETECTION SOLUTIONS: Empirical assessment of the effectiveness of solutions at identifying and blocking problematic respondents. Such assessment should include a description of methods and outcomes such as false positives and false negative rates.

OPERATIONAL DEFINITION: Clear, concise, and quantifiable definition of an outcome.

FALSE NEGATIVES: The proportion of problematic respondents who are incorrectly categorized as valid.

FALSE POSITIVES: The proportion of valid respondents who are incorrectly categorized as problematic.

GENERAL TERMS

SAMPLE: A smaller, manageable version of a larger group. It is a subset containing the characteristics of a larger population.

REPRESENTATIVENESS: The degree to which a sample accurately reflects the characteristics of a larger group.

PERSONAL INFORMATION: Any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means. Some examples that may be included here, but are not limited to, are information such as name, email address, phone number, street address, driver's license number, IP address, NPI for physicians or professional license numbers for any industry. This can be ever changing so look at GDPR for the latest. This is the industry term now being used in place of PII (personally identifiable information).

PRIVACY LAWS: The body of laws that deals with the regulating, storing, and using of personally identifiable information, personal healthcare information, and financial information of individuals, which can be collected by governments, public or private organizations, or other individuals. CPRA and GDPR are two examples of a privacy law.

OPEN ENDS: Questions designed for respondents to answer without a pre-populated answer set. Can be text or numeric.

QUALITATIVE RESEARCH: Research that uses open-ended descriptive responses, interviews, focus groups, and other forms of non-numerical information gathering, as opposed to closed-ended, more easily quantifiable survey methods.

QUANTITATIVE RESEARCH: Survey methodology that employs closed-ended response options that are easily numericized. Quantitative research is particularly vulnerable to fraud due to the ease with which closed-ended responses can be made by inattentive respondents and automated form fillers, and the relative difficulty in assessing the quality of closed-ended responses.

BENCHMARK COMPARISONS: Any project measure, result or response that is outside of acceptable ranges versus available benchmarks and may suggest data fraud or quality issues.

Goal: To collaborate with all stakeholders to continuously improve all quality areas.



HOW DO WE ADDRESS AND FIX THE PROBLEM?

Understand the sample source, delivery platforms and in-field methods to detect and deter fraud

It's important to gain an understanding of the respondents who participate in your survey research and how fraud is detected and mitigated throughout the survey process.

This includes the entire research chain, including study design, sample sourcing, programming and hosting, delivery platforms, and data analysis.

HERE ARE SOME CONSIDERATIONS:

- What type of sample is being used to source your study?
 - For example, is sample first-party research panel, loyalty and community panel, aggregated sample, river sample, affiliate traffic, etc?
- How is the sample sourced? Understand suppliers' acquisition strategy.
- What type of incentive program is utilized to reward respondents?
 - What are the tools and methods being used to detect and mitigate fraud (differentiate between proprietary and commercially available solutions)?
 - What techniques are used to verify respondents' identity?
 - Are behavioral fraud markers being tracked, and what types?
 - Is machine learning being used, and how is it applied?
 - Is digital fingerprinting being used to understand device inconsistencies?
 - How are BOTS being addressed?
 - How are duplicates addressed?

- Are links to and from the survey secure? This includes all the intermediary links from the panel to hosting and delivery platforms.
 - How is status being passed back?
 - Are the redirects and postbacks secured or can the link be manipulated?
- What quality checks are in place through the survey program in real time?
- What quality controls are utilized by the third-party delivery platform (i.e., marketplaces) used to field the study?
- How is a 360-loop applied in cases where suspicious or fraudulent respondents are identified by the researcher?
 - Are IDs in question shared with the supplier?
 - Are all parties in the ecosystem transparent and taking ownership of the issue?
 - Are reasons for the fraud explained and what actions are taken as a result?

- If only one in survey, then not lean heavy into this check
- Is straightlining likely or possible and should not count against a respondent?
- Sometimes helpful – a call for action in grids (please select option C or opposing statements)?
- Open Ends
 - Depends on
 - Importance of OE
 - Where is it in survey
 - How does all other data look?
 - How are we accounting for multiple OEs? When/how is it possible to read across open ends to check for contradictory responses?
 - Are we considering if respondents may be able to add detail needed?
- Red Herrings/Contradictions
 - Fact based vs past behavior vs opinion based
 - Placement: helpful if done early (in screener) as a way to remove prior to entering survey
 - Ghost brands tied to ownership
 - Low incidence checks and higher click rate across multiple

IN-SURVEY CHECKS

Quality measures and checks inserted into survey questionnaires are commonly used by researchers to evaluate the quality of online data for any given study. There are many different types of in-survey checks. The key to success is to identify the correct failure criteria for each check. It is also important to include multiple checks to identify a pattern of suspicious behavior.

COMMON THEMES FOR IN-SURVEY CHECKS:

- Speeding
 - Thresholds for survey length to flag
 - 1/4 vs 1/3 vs 1/2 median survey lengths
 - Are we accounting for different paths and survey methods?
 - Should we drill down for specific in-survey exercises?
- Straightlining
 - SL cleaning across multiple grids

- Survey design: are we placing any quality measurements? At least considering these:
 - IA Standards for respondent experience
 - Participate in the survey on all levels (programmer, field team, end client). How does the survey experience feel tied to each of these?
 - Survey length across all paths
 - # of Open Ends
 - Mobile compatible
 - Screener design: funnel approach vs direct to the point when panel is targeting
 - More rigor with audience (consumer, b2b and HC)
 - Tracking in-survey answers vs what they have answered previously (known panelist info)
 - Checking for red flags in terms of participant honesty – how to check they are who they say they are?
 - For guidelines and examples of in-survey questions, refer to document found [here](#).

WHAT'S NEXT - FUTURE WORK AHEAD

VALIDATIONS/QUALITY CHECKS SPECIFIC TO QUALITATIVE

Validations and quality checks are vital for quantitative research, but can and should be considered crucial for qualitative work, as well. Qualitative research has been shifting from face-to-face (which has an inherent quality check built in) to digital for about a decade. Covid, of course, accelerated that transition and pushed us further into a space where research hovers between quantitative and qualitative. That is, many databases that were historically only used for quant or qual are now being used for both.

While the data integrity and quality checks included here can be used for quant and qual, the Insights Association's DII Council recognizes the need for data quality checks and guidelines specifically for the nuances of qualitative research. With that goal in mind, IA's DII Council will add to the DII Toolkit by developing and collecting guidelines and resources unique to qualitative research. Look for the additions specific to qualitative research to be launched in Q1 of 2023.

In addition to the collection of information included here, survey quality and the respondent experience in survey research can greatly influence data quality and integrity. The Insights Association is excited to collaborate with ESOMAR on a project that will study factors that influence the respondent experience in survey research in the coming months.

