



# ESOMAR/GRBN GUIDELINE FOR ONLINE SAMPLE QUALITY

## ONLINE SAMPLE QUALITY GUIDELINE

ESOMAR, the World Association for Social, Opinion and Market Research, is the essential organisation for encouraging, advancing and elevating market research.

GRBN, the Global Research Business Network, connects 38 research associations and over 3500 research businesses on five continents.

## ONLINE SAMPLE QUALITY GUIDELINE

This guideline is drafted in English and the English text is the definitive version. The text may be copied, distributed and transmitted under the condition that appropriate attribution is made and the following notice is included "© 2015 ESOMAR and GRBN".

## CONTENTS

<b>1</b>	<b>INTRODUCTION AND SCOPE</b>	<b>5</b>
<b>2</b>	<b>DEFINITIONS</b>	<b>5</b>
<b>3</b>	<b>KEY REQUIREMENTS</b>	<b>7</b>
3.1	The claimed identity of each research participant should be validated.	7
3.2	Providers must ensure that no research participant completes the same survey more than once	9
3.3	Research participant engagement should be measured and reported on	10
3.4	The identity and personal data of research participants must be protected	10
3.5	Take special care with children and young people	11
3.6	Employ good online questionnaire design practices	11
3.7	Online sample providers must be transparent with researchers and clients about sample sources, the sampling process, and its outcomes	11
3.8	Researchers using online samples must be transparent with clients.	12
3.9	Passive Data collection	12
3.10	Conform to all relevant laws, regulations, and industry codes of conduct	13
<b>4</b>	<b>SPECIAL CONSIDERATIONS</b>	<b>13</b>
4.1	Category and other types of exclusions	13
4.2	Mobile	13
4.3	Survey and sampling automation	14
<b>5</b>	<b>PROJECT TEAM</b>	<b>14</b>
<b>6</b>	<b>REFERENCES</b>	<b>14</b>

## 1 INTRODUCTION AND SCOPE

With the emergence and general acceptance of online samples for market, opinion, and social research comes a responsibility to measure and ensure the quality of research results using such samples. A number of concerns about sample quality have been raised across the industry. They include concerns about:

- Professional research participants who through various means try to maximize their survey opportunities
- Inattentive if not deliberately untruthful research participants
- The potential for duplicate research participants as online sample providers broaden their sources in search of more diverse samples and low incidence populations
- Representativeness, that is, the degree to which research results accurately reflect the target population, however defined

This Guideline responds to those concerns by providing guidance on the operational requirements for the provision of online samples for market, opinion, and social research. It sets out methods to be used by online sample providers, buyers, and end clients to ensure that a sample meets widely accepted quality criteria. It is recommended reading for all stakeholders in the research process, from survey designers to data users.

It is meant to apply to all types of online samples, including those recruited from panels, from social media, and by web intercept methods. It is not meant to cover client-supplied samples such as lists of customers. Nor is it meant to cover samples recruited for qualitative research such as focus groups or one-on-one interviews since direct interaction between the researcher and research participant provides opportunities for quality assurance that is generally lacking in self-administered quantitative research. Nonetheless, researchers may find some of its suggested practices useful when working with those sample sources.

This guideline is not intended to substitute for a thorough reading and understanding of the [ICC/ESOMAR International Code on Market and Social Research](#), which has been adopted by over 60 local associations worldwide, or the individual codes of the 38 associations that comprise the [GRBN](#). Rather, it is intended to be an interpretation of the foundational principles of those codes in the context of online research. It also draws on a number of sources for its basic principles including ISO 20252 – Market, opinion, and social research and ISO 26362 – Access panels in market, opinion, and social research. It recommends full transparency with clients and encourages the use of a common set of terms and definitions. It offers a set of suggested practices for online sample quality, although it is not meant as a substitute for either ISO 20252 or ISO 26362, both of which require an external audit of a sample provider to ensure compliance. Many buyers of online samples may wish to limit their sample purchases to ISO-certified providers because of the added assurance that an external audit provides.

This Guideline also is not intended to be inclusive of all factors that might impact online research quality. Research on research in this area continues to evolve and there are many unsettled issues. However, there is a limited set of issues that have come to the fore despite some ongoing controversy about their impact on data quality. They are:

- Research participant validation
- Survey fraud prevention
- Survey engagement
- Category and other types of exclusions
- Sampling (including sample selection, sample blending, weighting, survey routers, profiling and screening).

Throughout this document the word “must” is used to identify mandatory requirements. We use the word “must” when describing a principle or practice that researchers are obliged to follow in order to comply with existing codes and standards such as those cited above. The word “should” is used when describing implementation and its usage is meant to recognize that researchers may choose to implement a principle or practice in different ways depending on the design of their research.

## 2 DEFINITIONS

**Access panel** Database of potential research participants who declare that they will cooperate with future data collection requests if selected.

**Completion rate** The number of research participants who fully complete a survey divided by the number of research participants who begin the survey or at least view the first or landing page.

**Consent** The freely given and informed agreement by a person to the collection and processing of his/her personal data. In market, social, and opinion research, this consent is based on the research participant having been provided with clear information about the nature of the data being collected, the purpose for which those data will be used and the identity of the person or organisation holding the personal data. The research participant may withdraw his or her consent at any time. A record of the agreement and how it was obtained must be kept.

**Cookies** Cookies are text files containing small amounts of information, which are downloaded to a computer, mobile device or other device when a user visits a website. Cookies are then sent back to the originating website on each subsequent visit, or to another website that recognizes that cookie.

Cookies are useful because they allow a website to recognize a user's device and make website navigation more efficient, remembering user preferences, and generally improving the user experience. They can also help to ensure that offers a user gets online are more relevant to them and their interests.

**De-duplication** For access panels, a process to remove individuals who are registered more than once on the same access panel so that they are entered only once.

For survey samples, a process to remove individuals who complete, or attempt to complete, the same survey more than once, whether because of having received multiple invitations or as a deliberate attempt at fraud.

**Device ID** A technology-enabled system that establishes a set of configuration data about a research participant's device (computer, smartphone, etc.), which can be used to create a machine or device fingerprint. Such systems assume the "machine fingerprint" uniquely identifies a device using settings and characteristics associated with an individual device or, potentially, an individual user account. Device ID systems apply to computers, mobile devices, and other devices that may be used to access the Internet where surveys can be completed.

Note: Device ID is also referred to as "digital fingerprint," a "machine fingerprint" or a "machine ID".

**Duplication** A situation where a research participant attempts to complete, or completes, a given survey more than once. This can occur, for example, when a panelist or research participant is a member of more than one panel or sample source (panel or sample source overlap) and is selected to participate in a survey that is split across sample sources and fails to recall previously participating in a given survey.

**Fraudulent responding** A situation where a research participant knowingly responds to survey questions in a dishonest way. A common example is the practice of giving false answers to screening questions in order to qualify for a survey.

**Online sample provider** A service provider responsible for the provision and management of online samples from relevant sources including panels, web intercept based sources (including river sample sources), email lists, etc. Online sample providers may offer some or all of the services below:

- Access panels
- Sample brokerage (the practice of purchasing and reselling sample)
- Sample aggregation (the practice of combining or aggregating multiple sample sources)
- Routing technology or access to routing technology

Online sample providers may provide other research or research-related services.

**Participation rate** Number of panel members who have provided a usable response divided by the total number of initial personal invitations requesting members to participate, in the case of an access panel that relies exclusively on such invitations. Defining participation rate for "river" and other, non-email based approaches is more complicated, with no approach as yet identified as a best practice.

Note: A usable response is one where the research participant has provided answers to all the questions required by the survey design. Where it is possible to determine undelivered invitations (e.g. returned to sender owing to a full email inbox, incorrect postal or email address, or invalid phone number), then these should be taken into account when calculating the participation rate. The number of panel members who did

not receive an invitation would then be subtracted from the total number of panel members invited to participate.

**Passive data collection** Research methods that collect data without the traditional use of survey questions.

**Passive validation methods** Internet-based methods used to measure research participant characteristics and panelist validation. These methods may include tracking research participant and panelist website visits, the specific pages they visit, and the links they click, then using that information to create a profile. This profile then is used to provide validation information for research participants and panelists.

**Research participant** Anyone taking part in a research project. In many cases their personal data are collected during that research project, by an active interview or by passive means.

**Representativeness** The degree to which a sample reflects the target population being studied. A representative sample is one in which the distribution of important characteristics is approximately the same as in the target population. The definition of “important characteristics” generally is a function of the survey topic(s). Representativeness may be achieved through sample selection, post survey adjustment (weighting), or a combination of the two.

**River sample** A non-panel sample recruited via the placement of online ads, offers, or other similar real-time invitation mechanism.

Note: River sample may also be referred to as a web intercept, real time sampling, and dynamically sourced sampling.

**Router** An online software system that screens incoming research participants and then uses those screening results to assign research participants to one of multiple available surveys. A router can also be used to offer research participants additional screeners and surveys after a screener qualification failure or survey completion. Routers are generally defined as *serial* or *parallel*.

A *serial* router generally utilizes a routing process whereby a research participant is screened sequentially for the available studies in the routing environment. Upon qualification for a study, the research participant is often immediately routed to that particular survey. Upon disqualification during the screening process, the research participant is reallocated, that is, he or she is re-routed to another screener (associated with another study within the environment). The process repeats until the research participant meets the qualification for an active survey within the routing environment, or, based on business rules (e.g. length of time in the process, number of screening questions that the research participant has answered by clicking, number of screeners that the research participant has been administered), the process ceases and the research participant is thanked and terminated.

A *parallel* router generally utilizes a process in which a research participant is exposed to a set of pre-screening questions from all or a subset of the surveys in the routing environment, simultaneously, on the same page. After the research participant answers these pre-screening questions, he or she is assigned to one of the surveys for which the research participant appears to be qualified. At that point, more specific screening may be conducted before the research participant is deemed eligible to complete the survey.

**Sample** A subset of the target population from which data are to be collected.

**Sampling frame** A list of population elements or other appropriate sources from which a sample is drawn.

**Sample blending** The practice of combining multiple, heterogeneous sample sources with the aim of achieving a more consistent or more representative sample. This practice typically utilizes balancing techniques at sample selection and may utilize sample profiling, scoring, or matching techniques.

**Satisficing** A survey taking behavior in which the research participant gives less than full cognitive effort when answering survey questions. Example behaviors include choosing non-substantive response options; choosing socially desirable responses; straight-lining in matrix questions; and acquiescence, that is, the tendency to agree with any assertion regardless of content.

## 3 KEY REQUIREMENTS

### 3.1 The claimed identity of each research participant should be validated.

Researchers and clients have long shared the concern that some people claim false identities as a way to maximize their survey opportunities and subsequent rewards or incentives. As a consequence, it has become common practice in some sectors of the industry for online sample providers to validate the claimed

## ONLINE SAMPLE QUALITY GUIDELINE

identify of every research participant, whether at the registration stage for online panels or the individual survey stage for non-panel sources.

Recent research conducted by the Advertising Research Foundation (ARF) and others has questioned the advantage of identity validation in improving data quality, finding little or no significant differences in response patterns among those participants who are successfully validated, those who are not successfully validated, and those who refuse to provide the personal information needed for validation. Further, this research has found that identity validation might also result in reducing the pool of available participants by removing participants who are not fraudulent. Nonetheless, many clients continue to insist on some form of participant validation and ISO 26362 requires it, where possible. Researchers also should note that these findings apply to consumer samples and validation continues to be critical with business-to-business sample, especially samples of physicians and other licensed professionals.

The specific variables used in validation will vary depending on the validation method used, the sources available for validation, and the restrictions, if any, imposed by local laws and regulations. Frequently used variables include:

- Full name
- Postal address
- Telephone number
- Year/Date of Birth
- Email address.

If the variables listed above are not available, applicable law and regulation prohibits their use, or the local cultural/social context contradicts their use, the online sample provider may use other appropriate methods, including passive methods (provided that such passive methods are not prohibited by applicable law and regulation).

The available sources (which can include address validation service providers, information service providers, etc.) also will vary based on geography and the population being studied. The most appropriate sources are those having information on a very large proportion of the target population.

This guideline recognizes that samples drawn from non-panel sources such as river sampling and some forms of routing pose significant validation challenges, as do those drawn in many countries where some forms of external validation are prohibited by law. While passive methods may be possible, their effectiveness has yet to be demonstrated. In all cases, the specific sources and methods used, any difficulties encountered, and validation outcomes must be documented and shared with clients upon request.

An expanded set of variables may be required when dealing with specialized populations such as physicians or other professionals. This expanded list may include but is not be limited to:

- Full name
- Business postal address
- Business telephone number
- Business email
- Appropriate available professional identification numbers (if relevant and available)
- Professional specialty (if relevant and available).

The data sources used for validation also may vary based on such factors as the type of target research participant or the geographic area being studied. This guideline recognizes that data sources available and useful in one country may not be available and useful in other countries. It also recognizes that the data sources used for validation are seldom all inclusive and that techniques for automatically matching identifying information collected from survey research participants with such sources can result in false positives as well as false negatives. Therefore, multiple data sources should be used where they exist. Further, given these inherent uncertainties in validation outcomes, online sample providers are encouraged to develop outcome codes that express the level of certainty of the identity of each research participant (rather than the use of a simple binary indicator indicating success or failure). For example, a research participant might be matched on only a subset of the validation data collected.

The specific sources used for validation must be documented and provided to clients upon request.



Finally, sample providers should carefully review identity validation failures to determine those research participants that should be removed from a panel, survey or other research activity. As identify validation methods may result in both false positives and false negatives, sample providers must exercise careful judgment in this practice and be transparent with clients and data users.

### **3.2 Providers must ensure that no research participant completes the same survey more than once**

As sample providers increasingly use multiple sources (multiple panels, social networks, river samples, etc.) to develop their samples it becomes increasingly likely that the same research participant(s) may be invited to and possibly complete the same survey more than once. Duplicate research participants must be removed prior to analysis, either by the sample provider or the researcher.

#### 3.2.1 Device ID

One common method of de-duplication uses the Device ID from a research participant's computer or device. Device ID is often referred to as a "digital fingerprint," a "machine fingerprint" or a "machine ID." The Device ID is typically created using variables or parameters from a web browser and typically includes:

- Operating system
- Time zone
- Language
- Browser type
- Browser parameters
- Flash ID
- Cookie ID
- IP address

In addition, the Device ID technology used should be capable of supporting geo-location identification and both duplicate and proxy server identification where possible. However, researchers should recognize that IP address and geo-location data may not be completely accurate, as multiple computers connected to a company's network or the same Internet Service Provider may share a common IP address. In addition, research participants can use various techniques to mask IP address and geo-location data.

The use of Device ID has raised privacy concerns in some jurisdictions. Online sample providers and researchers must ensure that any use of the technology complies with local laws. The Device ID technology used must not access any personally identifiable information and only the Device ID itself can be transferred or stored in a database.

Unfortunately, de-duplication methods that rely on Device ID can be problematic. As with validation of identity, both false positives and false negatives are possible. The increased use of mobile devices to complete surveys means that a more limited set of browser parameters are available to construct a Device ID. As a result, online sample providers and researchers are encouraged to develop outcome codes that express the level of certainty that two or more research participants are duplicates (rather than the use of a simple binary indicator).

#### 3.2.2 Cookies and other similar objects

Online sample providers routinely use or cooperate with third parties that use cookies and other similar objects, including local shared objects (commonly referred to as "flash cookies"), web beacons (including transparent or clear gifs) for panels and surveys. The legitimate use of cookies and other similar objects include:

- Identification of panelists or research participants that are required for services requested by the user (to participate in panels and surveys)
- Validation and fraud prevention, including legitimate use in Device ID technologies
- Advertising evaluation research

When cookies and other similar objects are used in panels and surveys, online sample providers and researchers must comply with all applicable laws, regulations, and industry codes, including the separation

of research and marketing activities. In some jurisdictions, this includes obtaining panel member and research participant permission to place cookies and other similar objects on their devices for the first time. Research participants must be told what cookies and other similar objects are and why they are being used. This information must be presented in language that is easily understood so that panelists and research participants can make an informed choice about whether to give their permission.

### 3.2.3 Other Methods

Online sample providers may use alternatives to Device ID technology, cookies, and other similar objects if they accomplish the equivalent functions at the same level of accuracy and effectiveness. These methods include other technology solutions as well as process-based solutions. Any alternative method must comply with local laws and regulations.

Regardless, the method used must be fully documented, and the results of the de-duplication process provided to the client upon request.

### 3.3 Research participant engagement should be measured and reported on

There is widespread concern among clients that online surveys are especially vulnerable to questionable data supplied by research participants who do not give adequate level of thought to answering survey questions or deliberately provide fraudulent answers. It is important to identify these research participants so that their impact on a study's overall findings is minimized.

Research on research has identified a broad set of possible measures that may be used to identify inattentive research participants. These include, but are not limited to:

- Survey completion time
- Proportion of unanswered questions
- Extent of selection of non-substantive answers such as "Don't Know" or "Refused"
- Patterned responses in matrix or grid questions (e.g., straight lining, random responding, etc.)
- Detection of inconsistent responses such as asking both positively and negatively worded questions in the same attribute battery
- "Red herring" questions including multiple response items with low probability or fictitious answer categories
- Trap questions such as "Check the box on the far right" in a matrix
- Appropriate responses to open-ended questions.

The researcher designing the survey and the company hosting it generally share responsibility for identifying potential inattentive or fraudulent research participants. The appropriate division of responsibilities is a matter to be negotiated between the two parties. The use of multiple measures from the above list is strongly recommended along with a scoring system that aggregates these measures across each research participant with the scores being included on the research participant data record. Researchers and clients should work together to determine the specific measures to be used as well as the threshold in survey scores that determines which research participants, if any, are deleted. Sometimes these measures are calculated and these undesirable research participants dropped in real time within the survey while other times this is done in post-survey processing. The online sample provider should be prepared to replace any research participant whose data are deemed unacceptable by the client, as long as the criteria for "unacceptable" has been previously agreed to.

The measures used and the method of calculation of the overall score must be documented and shared with the client on request.

### 3.4 The identity and personal data of research participants must be protected

A large body of law, regulations, and industry ethical codes require that any and all data collected from research participants online must be kept securely and only used for market research purposes. No personally identifiable data may be shared with a client without the consent of the research participant, and when done so it must be in compliance with national and local laws, regulations, and industry codes. When consent is given to share data with a client, the responsibility to keep the data secure and protect the identity of research participants transfers to the client. The transfer of such data and the transfer of the associated responsibilities must be clearly stated and documented via written agreements. Where such data are

transferred across national borders they must be made in accordance with laws and regulations such as the EU Data Protection Directive, an approved program such as US-EU Safe Harbor or formal data transfer agreements.

### 3.5 Take special care with children and young people

Online sample providers must ensure that no child is selected for a research project unless a parent or legal guardian, or other person legally responsible for the child has given permission for that child to participate in the specific project for which he or she is sampled. The legal definition of a child varies substantially from jurisdiction to jurisdiction and the sample provider must comply with the laws in the jurisdiction in which the child lives. Where there is no specific national definition, those aged under 14 should be treated as “children” and those aged 14-17 as “young people.” These age ranges generally recognise the different stages of mental and psychological development. For further guidance consult the ESOMAR guideline, [Interviewing Children and Young People](#).

Buyers of online sample must take care to ensure that the appropriate permissions have been obtained prior to interviewing children.

### 3.6 Employ good online questionnaire design practices

Despite almost two decades of online research and a significant body of research on research about online questionnaire design there are few widely accepted best practices. For example:

- The longer the questionnaire, the more likely research participants will disengage and potentially jeopardize data quality. A number of studies have shown an increase in satisficing behaviors and even breakoffs after 18-20 minutes.
- Research also has shown that a repeated series of matrix or grid-style questions can result in straight-lining or other patterned responding.
- A phenomenon known as primacy, where questions with a large number of answer categories can result in research participants choosing responses from the top of the list more often than the bottom.

In general, the best questionnaire practices are those that result in interesting and easily understood well-designed questions presented in a logical order with an equally well-designed list of potential answer choices. It is in the best interest of all parties—sample providers, researchers, data users, and clients—that questionnaires follow industry best practices.

### 3.7 Online sample providers must be transparent with researchers and clients about sample sources, the sampling process, and its outcomes

If users of online samples are to have confidence that their sample is fit for purpose, then online sample provider must make available information about the sample development process. Depending on the specific sampling approach used this includes:

- A description of the sampling frame or sources from which the sample was drawn (including any subcontractors used), how it was constructed or acquired, and the target population it is meant to represent.
- The sampling method used to select potential research participants from the sampling frame or equivalent and the means employed to ensure that the sample represents the target population, including any quotas or sample blending methods used.
- The specific criteria used in sample selection, such as quotas or other filtering criteria.
- The type(s) of incentives offered to sample members.
- Where panels invitations are used, a count of the number of sample units drawn and solicited, the number of bounced emails (which may alternatively be counted and removed at the panel level), the number of partial interviews, and the number of full, completed interviews.
- Where a router or similar intercept method is used, a count of the number of the number of potential research participants screened, the specific criteria used, the number of research participants qualifying, the rules with respect to the number of surveys a research participant is exposed to. When use of a specific router design is known to produce some bias in research participant selection, it also must be documented.

In addition, providers and users of online samples have additional reporting responsibilities due to the variety of sampling methods used for online research. While some online panels are recruited using traditional probability-based methods, most are not. Recent innovations such as online routers and advances in dynamic sourcing cast a still wider net across the Internet to solicit volunteers to complete surveys. As a consequence, the vast majority of online samples are convenience samples that lack the necessary statistical properties assumed to be needed to accurately represent the intended target population, thereby creating the risk that a study's results may contain significant error.

One common practice has been to impose demographic quotas (primarily age and gender) in sample selection or in post survey adjustment. A number of studies have shown that these adjustments are sometimes insufficient and additional adjustments using attitudinal or behavioural variables correlated with the survey topic may be needed to improve accuracy.<sup>1</sup>

ISO 20252 – Market, opinion, and social research requires that researchers report to clients “the procedure used to select potential research participants from the sampling frame or equivalent and the means employed to ensure that the sample represents the target population.” It further requires that researchers describe their weighting and projection methods and provide an “assessment of how well the sample represents the target population and the associated implications for data quality.” Similarly, ISO 26362 – Access panels in market, opinion and social research requires that “the access panel provider shall agree with clients on the design and methods to be used to draw samples from access panels for surveys or other research purposes.” It further requires that the sampling methods used shall be reported to clients. ESOMAR and GRBN consider these ISO requirements to be best practices that all researchers should follow.

### **3.8 Researchers using online samples must be transparent with clients.**

The appropriate reporting standards for research projects using online samples are similar to those for research projects in general. The following should be provided routinely to clients:

- The sampling frame or equivalent, sources and sampling methods used
- The dates of fieldwork
- The average or median survey length
- The total number of interviews completed
- Any quotas used or other specifications used in sample selection
- The questionnaire and other relevant data collection documents
- A count of the number of survey research participants whose identity was successfully validated
- A description of any de-duplication methods used and the number of responses deleted as a result
- The measures of research participant engagement used and an account of any research participants removed or replaced because of poor survey behaviour
- Exclusion information
- Participation rates<sup>2</sup> (where possible) and methods used to calculate them
- Completion rate
- Whether all or part of the project was subcontracted and, if so, to what organisations.

### **3.9 Passive Data collection**

---

<sup>1</sup> For further discussion see, “Report of the Task Force on Non-probability Sampling,” available on the website of the American Association for Public Opinion Research ([www.aapor.org](http://www.aapor.org)). A summary of the report along with critiques from several experts in the field has been published in *The Journal of Survey Statistics and Methodology*, Volume 1, Number 2, November 2013.

<sup>2</sup> This guideline recognizes that increasing use of river sampling and routers makes calculation of participation rates difficult if not impossible. Until a generally accepted best practice emerges researchers should note this difficulty in their reports to clients.

Passive data collection methods often involve data considered to be personal data. Sources of passive data include web browsing data, app data, loyalty data, geo-location data, social media data, and data generated by/obtained from mobile devices. Much of this data can be combined with survey data.

All passive data collection methods utilized must comply with local laws and regulations. As with personal data, researchers in many jurisdictions will have to set out a clear legal basis for using and processing these data, including its use for sampling activities and obtaining the consent of the individuals concerned.

### **3.10 Conform to all relevant laws, regulations, and industry codes of conduct**

It is critical that both online sample providers and buyers be aware of and strictly adhere to all relevant regional, national, and local laws and regulations as well as any industry codes of conducts or cultural dispositions that may set a higher standard than what is legally required.

## **4 SPECIAL CONSIDERATIONS**

As research methods and processes continue to change and evolve, new challenges to sample quality emerge. Despite these changes, the responsibilities of researchers using online sampling methods remain the same.

### **4.1 Category and other types of exclusions**

In some studies researchers may wish to exclude some research participants based on their recent survey taking experience to protect against bias. Examples include research participants who recently completed a survey about a similar product, a similar method (e.g. pricing study or concept test), or simply a very large number of surveys. Given the variety of sample sources now being used it is difficult if not impossible for a sample provider to know the full history of every research participant in a given sample. When responding to an exclusion request, sample providers must disclose the specific method(s) being used and document their effectiveness.

### **4.2 Mobile**

The proliferation of mobile devices (feature phones, smartphones, and tablets) worldwide has created new opportunities to engage with potential research participants. From a sampling perspective, two issues have emerged.

#### **4.2.1 Unintended Mobile Research Participants**

There is substantial evidence that an increasing number of research participants are choosing to respond to online surveys using a mobile device. Current research on research suggests that, on average, 20% to 30% of research participants may respond using a smartphone or other mobile device. It is reasonable to assume that this phenomenon will grow over time. Further, research on research also has shown that these research participants are often demographically different from other research participants— younger, more ethnically diverse, more likely to be male than female, etc. At the same time, online surveys are not always optimized for the smaller screens of many mobile devices, sometimes leading to lower completion rates or biased survey responses.

The responses to this increasing use of mobile devices by participants generally fall into three categories: (1) not allowing participation on a mobile device; (2) allowing research participants to respond, but making no allowance for mobile device limitations; and (3) optimizing survey presentation for mobile, although best practices in this area remain undetermined.

All three of these approaches have the potential to create some form of bias. Therefore, researchers should include the following in their reports to clients:

- Whether research participants were allowed to respond using a mobile device
- The number of surveys completed, in whole or in part, on a mobile device
- The implications for the sample's representation of the target population.

#### **4.2.2 New sample sources for mobile market research**

A second use of mobile devices in research falls in a category that might be called "pure mobile research." This research category capitalizes on the unique capabilities and use patterns of these devices and includes "in the moment research," use of geolocation information, and new forms of passive data collection. These

## ONLINE SAMPLE QUALITY GUIDELINE

methods typically require samples of known mobile users. Many online sample providers have created mobile panels. Online sample providers using river sampling and other forms of dynamic sourcing have begun to screen for mobile users and have added uniquely mobile sources such as mobile advertising networks. In all cases, both sample providers and researchers should provide the same level of scrutiny and reporting that is described in section 3.8 of this guideline.

### 4.3 Survey and sampling automation

The growing demand for the automation of research services poses a new set of challenges for managing and ensuring sample quality. Key features of such offerings include automation of the sampling process (including online sample source exchanges) and fast turnaround. The sample quality safeguards described in this guideline are not always built into DIY research services and transparency may be lacking. Researchers considering the use of such services should demand transparency and evaluate their options carefully, just as they would when selecting a traditional online sample provider. The concepts presented in this guideline and other sources including, [28 Questions to Help Buyers of Online Samples](#), may be used as a good starting point when undertaking such an evaluation.

## 5 PROJECT TEAM

Reg Baker, Co-Chair, Consultant to the ESOMAR Professional Standards Committee, Marketing Research Institute International

Peter Milla, Co-Chair, Technical Consultant to CASRO, Peter Milla Consulting

Pete Cape, Survey Sampling International

Mike Cooke, GfK

Melanie Courtright, Research Now

George Terhanian, The NPD Group

## 6 REFERENCES

[28 Questions to Help Buyers of Online Samples](#)

[ESOMAR Guideline for Conducting Mobile Market Research](#)

[GRBN Mobile Research Guidelines](#)

*The Impact of Digital Fingerprinting and Data Verification on Data Quality*, Advertising Research Foundation's Fundamentals of Quality 2.0, Courtright and Pashupati, 2014

[Interviewing Children and Young People](#)

ISO 20252 – Market, Opinion, and Social Research

ISO 26362 – Access Panels in Market, Opinion, and Social Research