“Factors affecting data quality”

Internet-based surveys have steadily gained popularity with researchers because of their myriad advantages such as ability to reach a larger pool of potential participants within a shorter period of time. This is in addition to obvious reasons such as convenience, relative inexpensiveness and user-friendly features such as comfortable pace and enhanced sense of participant control.

Many methodological issues confront a prospective researcher while designing online questionnaires/surveys. A few of them and some corresponding suggestions for troubleshooting are outlined below:

Web or mailed questionnaire?—A meta-analysis of 39 studies[4] concluded that response rates to mail surveys are, in general, higher than web surveys. Interestingly, two important factors underlying this variation were the type of respondents and medium of follow-up reminders. While college students were more responsive to web surveys, physicians and laypersons were found to be more receptive to mail surveys. Further, follow-up reminders were more effective when given by mail, probably owing to greater personalization, than the web. Recently, a hybrid method called web-push surveys, wherein, initial and subsequent follow-up contacts are made by mail to request a response by web was found to be a parsimonious method to elicit responses.[5] However, results on combining web and mail surveys have not been consistent. Examination of different methods of using web portals for surveying consumers about their experience with medical care revealed that response rates were higher for mail questionnaires, compared with web-based protocols.[6] A study on physicians[7] found that initial mailing of questionnaire followed by web surveying non-responders increased response rates and enhanced the representativeness of the sample. For surveys with shorter time frame, the reverse method, namely initial web survey followed by mailing the questionnaire to non-responders was recommended, though key outcome variables did not differ between these data collection methods. The message appears to be that hybridization of surveys involving both web and mail surveys may be a resource-effective method to augment response rates, though the optimal method of combining them may differ based on respondent and survey characteristics.

How to enhance response rates?—Response rates to email surveys are highly variable and traditionally in the range of 25%–30%, especially without follow-up reminders or reinforcements.[8] Reasons for this include survey fatigue, competing demands and privacy concerns. These response rates are much lower than traditional response rates to telephonic surveys, which in turn are less than response rates to surveys using the face-to-face method.[9 10] However, with the use of multimode approaches, the response rate can be improved to as much as 60%–70%.[11] Substantial evidence exists on methods to enhance response rates in internet surveys. More than a decade ago, a Cochrane review suggested pre-notifying respondents, shortening questionnaires, incentivizing responses and multiple follow-up contacts to enhance response rates.[12]Subsequently, personalized invitations,[13] personalization of reminders,[14] dynamic strategies such as changes in the wording of reminders[15] and inclusion of Quick Response (QR) codes[16] have all been found to augment response rates.

How long is too long?—In academic surveys, the association between questionnaire length and response rates, particularly for web surveys, is weak and inconsistent.[17 18]Considering the attention span of an average adult, 20 min has been recommended as the maximum length of web surveys by market research experts.[19] However, it is difficult to be dogmatic about this as many non-survey factors, such as familiarity with the device and platform, may affect interest and attention span.[20] On current evidence, restricting survey length to below

13 min may offer the best balance between response burden and quality.[21]

What are the optimal frequency and number of reminders?—Sending reminders have been noted to positively influence survey response rates.[21] Response rates appear to improve till a maximum of three to four reminders, beyond which concerns about spamming the respondents increase.[14] Less is known about the optimal frequency of reminders. Blumenberg and others (2019) noted that sending reminders once every 15 days was associated with higher response rates, especially to the first two questionnaires in a series. Researchers should carefully weigh the efforts involved in sending multiple reminders against the risks of survey fatigue and lower engagement in later surveys, especially when multiple rounds of surveys are indicated.

How to minimize the non-representativeness of the sample? In any internet survey, the respondents are not selected through probability sampling and this may affect the usability of findings. A larger sample, too, will not solve this issue because information about non-responders is not available. One solution, though more resource-intensive, would be to randomly select the required number of respondents from a defined list (such as professional membership directories in a survey of professionals), seek their consent through telephone calls, collect basic sociodemographic information from those who decline to participate and send the questionnaire to only those who consent to participate, with a request not to forward the questionnaire to anyone else.[22] This way, we prevent snowballing which can also affect sample representativeness. Further, if the sociodemographic profile of those who declined matches with that of those who consented to participate, then we can conclude that responders are representative of the intended population. Of course, this strategy may only work when a list of potential respondents is available. To avoid missing out those with no internet access or low-frequency users, investigators may contact people, randomly selected from a list, through another mode (telephone, face to face or mail) and ask them to complete the survey. Further, allowing respondents to complete the survey in a variety of modes (web/mail/telephone) may be another strategy to avoid missing out on those with limited access to the web.

In conclusion, hybrid survey designs incorporating email surveys with web-based reminders probably lead to quicker and higher response rates. Weekly reminders (maximum of 3–4), would serve to enhance response rates. Questionnaire length is not associated with response rates, but limiting the survey to under 20 min may be a pragmatic consideration. Other suggestions to improve response rates that must be treated as preliminary due to limited evidence include sending personalized email reminders, modifying reminder contents over the survey life-cycle and creating QR codes that may seamlessly direct respondents to the survey.

CONSUMER RESEARCH BACKGROUND READING

"Digital Survey Design for Community Engagement: Structural Considerations for Maximum Response Validity." Bhattacharya, Wong, & Martinez (2022) *Journal of Marketing Research*

Examines how survey structure affects response depth in community development contexts.

The authors conducted experiments comparing survey structures to determine which formats elicit the most meaningful qualitative feedback. Their key findings include:

- Sequencing effects - Placing open-ended questions after related closed-ended questions produced 27% longer and more substantive responses than when open-ended questions appeared first.
- Response box design - Larger text entry fields with expandable features resulted in responses with greater detail and actionable suggestions compared to standard-sized boxes.
- Cognitive priming techniques - Using brief scenario-based prompts before open-ended questions improved response quality by helping respondents contextualize their thoughts.
- Question framing - Community-focused language (e.g., "How would this benefit our neighborhood?") versus individual-focused language (e.g., "How would this benefit you?") generated responses with broader stakeholder considerations.
- Project framing – Holistic language that reminds respondents of other types of stakeholders and uses with a vested interest in the project.
- Visual elements - Strategic use of relevant images and progress indicators increased respondent engagement and reduced dropout rates by approximately 18%.

Thoughtful structural design significantly enhances the usefulness of qualitative data gathered through online surveys for community planning purposes, particularly when engaging diverse populations with varying levels of digital literacy.

"The Efficacy of Mixed-Method Digital Surveys in Public Planning Projects." Peterson, Williams, & Lin (2023). *Journal of Public Administration Research and Theory*

A comprehensive framework for integrating quantitative and qualitative approaches in digital surveys for public planning projects.

The authors analyzed data from 37 community planning initiatives across diverse settings to evaluate the effectiveness of mixed-method online survey approaches. Their key findings include:

- Complementary data integration - Surveys combining structured rating questions with open-ended response fields yielded more actionable planning insights than either approach alone.
- Sequential design benefits - The researchers found that when quantitative questions preceded qualitative ones, they provided context that improved the specificity of open-ended responses.

- Demographic response patterns - Different demographic groups showed varying preferences for question types, with significant implications for inclusive community representation and understanding if sufficient representation is captured by the survey.
- Visualization techniques - Interactive data visualization tools embedded within the survey increased participant engagement and improved response quality among typically underrepresented groups.
- Mobile optimization impact - Mobile-optimized surveys showed a 34% higher completion rate from lower-income and minority respondents compared to desktop-only surveys.

The authors propose a practical framework they call "CARES" (Context, Accessibility, Responsiveness, Engagement, Synthesis) for designing effective mixed-method digital surveys. Their work emphasizes that strategically structured online surveys can overcome traditional participation barriers in community planning while generating statistically significant quantitative data and rich qualitative insights.

"Response Bias Mitigation in Digital Consumer Research: Strategies for Survey Design." Kim, Roberts, & Zhao (2022). *Journal of Consumer Psychology*

Analysis of response biases in online consumer surveys and evidence-based strategies to mitigate these issues.

The authors conducted a series of experiments with over 4,000 participants to identify factors affecting response quality in digital consumer research. Their key findings include:

- Social desirability effects - The research quantifies how online surveys reduce but don't eliminate social desirability bias compared to in-person methods. The authors recommend indirect questioning techniques and self-administration features to minimize this bias.
- Satisficing behaviors - The study identifies patterns of survey "rushing" and offers specific design elements to increase cognitive engagement, including attention checks that are contextually integrated rather than obvious.
- Anchoring effects - The researchers demonstrate how initial questions create reference points that influence subsequent responses and provide techniques for neutralizing these effects through randomization and buffer questions.
- Time-of-day impact - Their data reveals significant variations in response quality based on when participants complete surveys, with implications for scheduling and timeframe selection.
- Device-specific considerations - The authors document how response patterns differ between mobile and desktop users, offering concrete design modifications to ensure consistency across platforms.

The paper concludes with a practical framework for response bias assessment and provides a decision tree for researchers to determine which mitigation strategies are most appropriate based on research objectives and target populations.

"Question Wording Effects in Online Consumer Surveys: Experimental Evidence." Summary: Torres, Davis, & Chen (2022). International Journal of Market Research

Experimental evidence on how question wording significantly impacts consumer responses in online surveys.

The researchers conducted a series of controlled experiments with over 3,500 participants to examine how subtle variations in question phrasing affect response patterns. Their key findings include:

- Framing effects - The study demonstrates that positively framed questions ("How satisfied are you?") versus negatively framed questions ("How dissatisfied are you?") produced significantly different response distributions, with positive framing generating more favorable responses by an average of 12%.
- Specificity impact - Questions with specific reference points (e.g., "compared to your last purchase") yielded more consistent responses than generally worded questions (e.g., "how would you rate").
- Emotional language effects - The inclusion of emotionally charged words (e.g., "love" vs. "like") systematically shifted response patterns, particularly in product evaluation contexts.
- Scale labeling influence - The experiments revealed that different verbal anchors on rating scales (e.g., "extremely satisfied" vs. "very satisfied") produced measurably different response distributions despite identical numerical scales.
- Question order interaction - The authors documented how wording effects were amplified or diminished based on question sequencing, with significant implications for survey structure.

The researchers provide practical guidelines for question formulation to improve measurement validity in consumer research, emphasizing the need for consistent, neutral language and pre-testing multiple wording options to identify potential bias sources.

"Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, Updated Edition." Dillman, Smyth, & Christian (2023). *Survey Research Methods*

A comprehensive framework for designing and implementing effective digital surveys within mixed-method research contexts.The authors present their refined "Tailored Design Method" based on decades of empirical research and technological advancements. Key elements include:

- Unified design principles - The researchers demonstrate how core survey methodology principles must be consistently applied across digital platforms while adapting to medium-specific constraints and opportunities.
- Device-responsive design - The work presents extensive evidence showing how optimizing surveys for multiple devices significantly improves completion rates and response quality, with specific technical guidelines.
- Trust development framework - The authors identify specific design elements that establish respondent trust in digital environments, a factor they demonstrate is critical for honest, thoughtful responses.

- Mixed-mode integration - The research provides protocols for seamlessly integrating online surveys with other methodologies (phone, mail, in-person) to maximize response rates while maintaining measurement consistency.
    - Sequential Implementation Strategy The authors advocate for a carefully sequenced approach that starts with less expensive methods (typically digital) before progressing to more resource-intensive modes for non-respondents. Their research demonstrates that this sequential approach can increase overall response rates by 15-30% compared to single-mode approaches while controlling costs.
    - Mode-Specific Question Adaptation One of the most valuable contributions is their evidence-based guidelines for adapting question formats across modes while maintaining measurement equivalence. For example:
    - Visual scales that work well online need specific verbal anchors when translated to telephone surveys
    - Complex grid questions effective on desktop devices need restructuring for mobile and verbal administration
    - Unified Contact Protocol The researchers outline a systematic contact strategy that coordinates outreach across multiple channels (email, text, mail, phone) with specific timing recommendations based on experimental evidence showing optimal intervals between contact attempts.
    - Response Integration Methodology The work addresses the critical issue of combining data collected through different modes by providing statistical procedures to identify and account for mode effects. This includes techniques for detecting mode-related measurement differences and analytical approaches to adjust for these differences.
    - Demographic Considerations The authors present extensive research on how different demographic groups respond to various survey modes, allowing researchers to strategically select modes that will effectively reach underrepresented populations and adjust sampling approaches accordingly.
    - Their comprehensive approach treats mixed-mode design as a holistic system rather than simply offering multiple response options, with particular attention to maintaining measurement consistency while leveraging the strengths of each mode.

- Error reduction strategies - The work details evidence-based techniques for minimizing four fundamental sources of survey error: coverage, sampling, nonresponse, and measurement.

1. Coverage Error Reduction Coverage error occurs when your sampling frame doesn't represent your target population. Their strategies include:

    - Using mixed-mode approaches to reach digitally underserved populations
    - Implementing device-detection technology that redirects respondents to optimized survey versions
    - Developing mobile-first designs for populations with higher smartphone than computer access
    - Creating "low-bandwidth" versions for areas with limited internet infrastructure

2. Sampling Error Reduction Sampling error represents the deviation between sample and population values. The authors recommend:

- Probability-based panel recruitment for online surveys rather than convenience sampling
- Stratified sampling techniques optimized for digital implementation
- Sample size determination protocols that account for mode-specific response patterns
- Weighting procedures specifically calibrated for digital survey contexts

3. Nonresponse Error Reduction Nonresponse error occurs when people who don't respond differ from those who do. Their research advocates:

- Evidence-based subject lines for email invitations (their experiments show specific approaches increasing open rates by up to 25%)
- Strategic timing of survey distribution based on device usage patterns
- Personalized reminder systems tailored to respondent characteristics
- Visual progress indicators designed to reduce abandonment at specific survey points
- Incentive structures optimized for different demographic groups

4. Measurement Error Reduction Measurement error involves inaccuracies in responses themselves. The authors' digital-specific strategies include:

- Visual design principles that reduce satisficing behaviors
- Input validation techniques that catch impossible or inconsistent responses
- Optimized question presentation for various screen sizes
- Interactive elements that improve respondent engagement with complex questions
- Format consistency principles to minimize context effects

Effective digital surveys require implementation of established methodological principles rather than technology-driven approaches.

Design with data in mind

What are the project’s objectives? Before crafting survey questions, define what type of data will achieve those objectives. Determining whether the project requires qualitative or quantitative data will shape the survey design process and ensure the information is actionable.

Conducting Research on the Internet: Online Survey Design, Development and Implementation Guidelines

Article *in* International Journal of Human-Computer Interaction · January 2003

CITATIONS
122

READS
62,504

4 authors, including:

Blair Nonnecke
University of Guelph
53 PUBLICATIONS **3,919** CITATIONS
SEE PROFILE

Jennifer Preece
University of Maryland, College Park
123 PUBLICATIONS **21,037** CITATIONS
SEE PROFILE

DRAFT: Andrews, D., Nonnecke, B., Preece, J. (2003) Electronic survey methodology: A case study in reaching hard to involve Internet Users. *International Journal of Human-Computer Interaction*. 16, 2, 185-210.

Conducting Research on the Internet: Online Survey Design, Development and Implementation Guidelines

Dorine Andrews, D.C.D.
Research Professor,
Communication, Culture and Technology Program
Georgetown University
Washington, DC
dca4@georgetown.edu

Blair Nonnecke, Ph.D.
Assistant Professor
Computing & Information Science
University of Guelph
Guelph, Ontario
rbn@acm.org

Jennifer Preece, Ph.D.
Professor and Chair, Information Systems Management Dept.
University of Maryland Baltimore County
Baltimore, Maryland
preece@umbc.edu

Abstract

Using the Internet to conduct quantitative research presents challenges not found in conventional research. Some of our knowledge concerning the effective design and use of paper-based surveys does translate into electronic formats. However, electronic surveys have distinctive technological, demographic and response characteristics that affect how they should be designed, when they can be used and how they can be implemented. Survey design, subject privacy and confidentiality, sampling and subject solicitation, distribution methods and response rates and survey piloting are critical methodological components that must be addressed in order to conduct sound online research. This paper focuses on those distinctive characteristics. It reviews the current literature on the subject of electronic surveys and presents guidelines for designing, developing and implementing them, particularly web-based surveys. This paper argues that Web-based surveys are superior to email surveys in many aspects, but that email combined, perhaps with offline media, is an excellent vehicle for inviting individuals to participate in Web-based surveys. The application of these guidelines are demonstrated through the authors' current research involving defining the nature of "non-public participation" (commonly referred to as lurking) in online discussion groups. Guidelines do not eliminate the many "trade-off" decisions required in the use of online surveys.

Keywords: Electronic surveys, web-based surveys, email surveys, methodology, online community, research techniques, Internet research.

Introduction

It is well recognized in the behavioral sciences that surveys are not perfect vehicles for collecting data because surveys require subjects to recall past behavior [Schwarz, 1999]. Some social scientists contend that observation captures behavior more accurately [Bernard, et al., 1981; Bernard, et al., 1983] and there is ample data to support their position. For example, online consumers overestimate the amount they purchase online by 55% [Comscore, 2001]. Others suggest that the survey questions bias subject judgements and answers [Schwarz, 1999]. One alternative, many contend, is to collect behavioral data using multiple approaches [Sudweeks & Simoff in Jones, 1999; Rogers, 1987]. Observations, focus groups, individual interviews, email, Web-based, postal, and

random digital dial telephone surveys can be used in combination to improve results quality [Smith, 1997] and sample representativeness [Yun & Trumbo, 2000; Swoboda, et al., 1997]. For example, Yun & Trumbo (2000) achieved a 72% return rate within a one-month period by combining postal, email and Web-based survey forms.

However, research costs, access to subjects, the scope of the research and the nature of behavior under study may make it impractical or financially unfeasible to use more than one data collection approach. Electronic surveys provide the ability to conduct large-scale data collection by others than organizations at the centers of power in society [Couper, 2000]. The technology provides an inexpensive mechanism for conducting surveys online instead of through the postal mail [Sheehan & Hoy, 1999; Weible & Wallace, 1998] and one in which costs per response decrease instead of increase significantly as sample size increases [Watt, 1999]. Electronic surveys are becoming increasingly common [Lazar, J & Preece, J., 1999], and research comparing electronic vs. postal surveys is starting to confirm that electronic survey content results may be no different than postal survey content results, yet provide strong advantages of speedy distribution and response cycles [Yun & Trumbo, 2000; Swoboda, et al., 1997].

Some of our knowledge concerning the effective design and use of paper-based surveys does translate into electronic formats. However, electronic surveys have distinctive technological, demographic and response rate characteristics that affect how they should be designed, when they can be used and how they can be implemented [Sohn, 2001]. This paper focuses on those distinctive characteristics.

Two forms of electronic surveys have emerged in the last fifteen years. The first, asynchronous email survey dates back to 1986 [Kiesler & Sproull, 1986]. The second, synchronous Web-based survey, started about 1994 [Kehoe & Pitkow, 1996]. There are several fundamental differences between email and Web-based surveys. First is database technology. Web-based surveys provide the ability to automatically verify and store survey responses using database technology and an HTML (hypertext markup language) user interface. Email surveys and responses, whether embedded directly within an email message or attached as a word processed document, must be manually transferred and entered into storage. Second, email is a "push" technology that allows researchers to directly communicate with prospective respondents. Web-based surveys do not provide this affordance of direct communication. This paper argues that Web-based surveys are superior to email surveys in many aspects, but that email combined, perhaps with offline media, is an excellent vehicle for inviting individuals to participate in Web-based surveys. Five methodological components of online survey design and implementation are critical to

successful Web-based surveys. These are (1) survey design, (2) subject privacy and confidentiality, (3) sampling and subject selection, (4) distribution and response management, and, (5) survey piloting.

Survey Design Guidelines

Technically, electronic surveys should be designed to (1) support multiple platforms and browsers [Yun & Trumbo, 2000], (2) prevent multiple submissions [Yun & Trumbo, 2000], (3) have the ability to present questions in a logical or adaptive manner, if needed [Kehoe & Pitkow, 1996], (4) provide multiple opportunities for saving the work in long questionnaires (e.g., over 50 questions) [Smith, 1997], (5) collect both quantified selection option answers and narrative type question answers [Yun & Trumbo, 2000], and, (6) provide feedback "thank-you" upon completion of the survey [Smith, 1997].

Email survey research has established that email surveys meet some, but not all, of the electronic survey criteria cited above. The format of email survey can accommodate the principles of paper questionnaire design [Oppenheim, 1992; Dillman, 2000; Preece et al, 2002]. These principles include the development of question scales and multiple choice answers from qualitative exploratory interview data, elimination of question bias through proper wording, and the use of clear, unambiguous and concise wording. Like postal surveys, successful email surveys have been shown to include: informed consent information, rating definitions and examples, rating scale formats such as Likert type, semantic differential scales and nominal scales, and a set of demographic items [Preece et al, 2002; Witmer et al, 1999]. In addition, open-ended questions can be successfully accommodated in email surveys. Respondents were found to write lengthier and more self-disclosing comments than they do on mail surveys [Schaefer & Dillman, 1998; Bachmann & Elfrink, 1996; Kielser & Sproull, 1986; Loke & Gilbert, 1995].

Email also affords the technical ability to track whether the delivered email survey was opened, responded to or/and deleted as well as if the survey was undeliverable [Paolo et al., 2000]. However, email surveys have significant technical drawbacks. They can be altered by the survey takers themselves [Witmer et al., 1999]. There is no way to prevent someone from changing, eliminating or adding questions to the survey. Email surveys have also been found to be confusing to complete by respondents [Sheehan & Hoy, 1999]. This may be caused by the fact that email survey completion is dependent upon the email software if the survey is included in as part of the email, or on the word processing software if the survey is attached as a document. How respondents enter the answers to the survey question may vary because of this. Some respondents may not know how to manipulate the survey text to

enter the responses correctly. In other words, the researcher does not have control over how the questions are displayed by software and how responses are entered into the email survey text.

Like email surveys, Web-based surveys have the advantage of low cost and quick distribution. Additionally, Web-based surveys provide the ability to transfer survey responses directly into a database, eliminating transcription errors and preventing survey alteration by the survey respondent. Initially technical issues inhibited the use of Web-based surveys, but new software and Internet related technology appear to be mitigating many of the technical limitations [Smith, 1997; Kehoe & Pitkow, 1996; McCoy & Marks, 2001]. Software applications such as Cold Fusion [McCoy & Marks, 2001] and software applications such as "Survey Wiz" and "FactorWiz" [Birnbaum, 2000] eliminate many of the construction and administration challenges of Web-based surveys.

The principles of paper design apply to Web-based surveys as with email surveys. Although questionnaire screen design is more complex because it must be developed in HTML and supporting scripting and database languages, Web-based surveys provide additional format and response control [Preece et al 2001; Stanton, 1998]. For example, radio buttons prevent multiple answers when only one is called for. Both coded and open-ended questions can be accommodated in Web-surveys. In a study using a Web-based survey where open ended questions were located after a set of coded questions, over 70% of the respondents provided additional information and explanations through the open ended question opportunity [Andrews et al, 2001]. However, it appears that attrition rates increase when using many open-ended questions requiring multiple items in the answers [Crawford et al., 2001] or when using questions that are arranged in tables on Web-based surveys [Knapp & Heidingsfelder, 1999]. This contradiction in attrition may be the result of question placement and whether the questions are optional or required.

The Web-based survey designer has a wide range of textual options, format control and graphics sophistication not attainable with email surveys. The advantages include links, clicks, defaults and menus [Preece et al, 2002]. Links provide the ability to directly reference definitions or examples at multiple points in the survey. Clicks eliminate the need for textual data entry for all coded questions. Defaults, hidden or displayed, reduce non-response to questions. Menus, drop-down or displayed, provide an economical way to display many response options without cluttering the survey screen. Additionally images, animation and color enhance survey presentation [Yun & Trumbo, 2000] but have disadvantages of increasing download time and may also affect the answers

subjects do or do not provide [Couper et al., 2001]. In addition, the survey designer cannot control the survey presentation as he/she would in a paper survey. Browser settings, user preferences and variations in hardware put the user in control [Couper, 2000]. Such variation and resulting poor design from the misapplication of Web-based technical capabilities increase the likelihood of response error and defeat Web-based survey advantages. Dillman et al. [1998] found that surveys with multiple or graphic designs that do not make clear what the respondent is to do resulted in higher attrition (drop out) rates than those surveys using more straightforward, plain designs. Dillman [2000] also warns that poorly designed Web-based surveys encourage novice Web-users to break off the survey process, making them less effective than email or postal surveys.

Table 1: Electronic Survey Design Evaluation

Design Items	Email	Web-based
• Supports multiple platforms and browsers	Yes	Yes
• Controls for browser settings	No	No
• Prevents multiple submissions automatically	No	Yes
• Presents questions in a logical or adaptive manner, e.g., provides control of when and how questions displayed	No	Yes
• Allows saving responses before completion	No	Yes
• Collects open-ended or quantified-option responses	Yes	Yes
• Provides automatic feedback with completion	No	Yes
• Can apply paper questionnaire design principles	Yes	Yes
• Provide automatic transfer of responses to database	No	Yes
• Prevents survey alteration	No	Yes
• Provides response control and economical displays	No	Yes
• Provides researcher control over question presentation	No	Some
• Provides for animation, sound, graphics options, etc.	No	Yes
• Does not requires familiarity with survey software	No	Yes
• Rapid display to respondent	Yes	Depends
• Technical ability to track response source	Yes	No
• Technically easy to design and implement	Yes	No

In summary (Table 1), Web-based surveys present many more design options than email surveys and provide researchers with increased control over respondent use of the survey. However, Web-based surveys are more challenging to design and more technically difficult to implement because of these options.

Subject Privacy and Confidentiality Guidelines

Online researchers may commit multiple violations of individual and online community privacy that can be more intense than those found in conventional survey methods [Cho & LaRose, 1999]. They define physical (unsolicited requests), informational (personal information control), psychological (personal choice control), and interactional (relationship control) privacy infringements that manifest themselves in both email and Web-based surveys. For example, in one study, email pre-notification and follow-up procedures for a survey were found to invade the individual's physical computer space. Receivers considered the email to be rude, unsolicited "spam"

[Schillewaert, et al., 1998; Swoboda et al., 1997]. Surveys violate informational privacy when they do not allow control over the conditions of release, use, retention, and disposal of personal data. Privacy is also breached when sampling lists are culled from non-public discussion groups and listservers.

The flexibility of the Internet and the ease with which false identities are created on the Internet exacerbate trust and confidentiality issues according to Cho & LaRose and can make survey results unreliable. Personal choice and voluntary prudence of others protect psychological privacy [Burgoon, et al., 1989] by allowing individuals to evaluate whether or not they wish to participate in the survey. Cho & LaRose [1999, p. 427] suggest online surveys that "rely on the trolling method to develop sampling frames or that contacting respondents through online communities raise the concern." This sampling technique, commonly called convienience sampling because large numbers of potential respondents are gathered without regard to demographic characteristics, is an attempt to reduce bias by getting large numbers of people to participate. As a result, researchers have observed uninhibited displays of emotional and psychological states that invade individual privacy causing large numbers to refuse to participate in the unsolicited surveys. Also, Cho & LaRose [1999] suggest that if researchers better manage interactional privacy, the willingness to disclose information increases. For example, they conclude that online surveys with little author reputation, authority and professionalism breach this interactional privacy as does entering communities because these communities are often sources for emotional support.

Therefore, Cho & LaRose make recommendations for mitigating privacy issues that may help build a trusting, quality relationship between the survey subjects and the researchers. These include (1) separate the invitation to participate (e.g. consent) from the survey questionnaire, (2) offer e-incentives as a trade-off for the intrusion, (3) collect data through Web pages, (4) provide multiple response options, (5) use remailers, entities that disguise actual email addresses with pseudonym addresses to prevent tracability back to the original author, to ensure anonymity in email surveys, (6) do not troll through observation, (7) do not use cookies, (8) do not use links from personalized sites, (9) provide disclosures, (10) certify privacy through 3rd parties, (11) use credible domains, (12) use encryption for extremely sensitive material, (13) use hypertext links for long disclosures, (14) disclose sampling procedures, (15) obtain consent of online community leaders to obtain email addresses and even have leaders provide the message, and, (16) post survey results or summaries of results. In addition, establishing credibility quickly in subject lines and opening statements would appear to be critical in managing privacy concerns

in email or Web-based surveys. Using "opt-in" email lists provided by email list brokers for market research is also recommended [Yun & Trumbo, 2000].

In some survey situations, subject identification is critical to longitudinal studies where the same subjects are surveyed multiple times. The use of self-selected user ids and the choice of "rather not say" for sensitive questions has been successful in providing some sense of privacy [Kehoe & Pitkow, 1996]. Some researchers suggest that lack of anonymity may not affect response rates [Couper et al., 1999] while others suggest anonymity is important to response rates [Kiesler & Sproull, 1986]. These conflicting findings may be the result of subject matter differences. For these reasons, some researchers guarantee confidentiality (i.e. no one will see your personal data or know you were a subject in the study), but not anonymity (i.e. the researchers will know who you are) and recommend explaining the method for keeping the confidentiality to the survey takers [Sheehan & Hoy, 1999].

In summary (Table 2), both email and Web-based surveys present issues of privacy and confidentiality infringement that must be addressed if researchers wish to collect personal information from respondents to ensure qualification to complete the survey. Collecting data through Web-based surveys provides improved opportunities to protect respondent confidentiality over email. Personal data can be stripped from the responses and manual transcription of completed survey responses is not required. However, building sufficient trust to get people to

Table 2: Subject Privacy & Confidentiality Evaluation

Privacy & Confidentiality Items	Email	Web-based
1. Respondent can designate conditions of release, use, retention and disposal of personal data	Yes	Yes
2. Do not sample from non-public email lists, discussion groups and listservs	Yes	Yes
3. Send invitations and surveys separately	Yes	Yes
4. Offer e-incentives	Yes	Yes
5. Collect data through web pages	No	Yes
6. Provide multiple response options	Yes	Yes
7. Use "remailers" to ensure anonymity	Yes	Not needed
8. Do not troll through observation	Yes	Not needed
9. Do not use "cookies"	Not needed	Yes
10. Do not use links from personalized sites	Not needed	Yes
11. Provide disclosures	Yes	Yes
12. Provide 3rd party privacy certification	Yes	Yes
13. Use credible domains	Not needed	Yes
14. Use encryption for sensitive material	Yes	Yes
15. Use hypertext links for long disclosures	No	Yes
16. Disclose sampling procedures	Yes	Yes
17. Community leader consent for member email addresses can be obtained	Yes	Yes
18. Provide survey results to respondents	Yes	Yes
19. Use self-selected user ids, passwords (option)	No	Yes
20. Provide "rather not say" response option for sensitive questions	Yes	Yes

participate requires multiple disclosure, design and distribution strategies on the part of the researchers and the enlistment of third party guarantors.

Sampling and Subject Selection Guidelines

Couper [2000] provides a topology for Web-based surveys based upon sampling technique. This topology includes (1) non-probability methods of sample selection entertainment, self-selection, and volunteer panels of Internet users, and, (2) the probability-based sample selection methods of intercept, list-based high-coverage, mixed-mode design with choice of completion method, pre-recruited panels of Internet users, and probability samples of full populations. Researchers are concerned with all except the first type -- Web-based surveys as entertainment. Couper [2000] defines each type and its advantages and disadvantages. Self-selection Web-based surveys post invitations to participate at multiple online locations. There is no attempt to statistically sample the online population, although some claims for scientific validity are sometimes made. Volunteer panels of Internet users are fast growing in industry of late. Subjects are selected for the panel by submitting demographic information at a portal, then are asked to participate by invitation-only to a survey. Although researchers have more information about the subjects, the base of this approach is still self-selection, not a sound statistical sampling approach. Probability-based methods begin with knowledge of a sampling frame and with information on the process of recruitment that permits measurement of sources of non-response, which can inform design-based adjustment approaches. These types of surveys either restrict the sample to those with Web access, thereby defining the population or use a mixed mode to reach a broader sample of the population. Intercept surveys target visitors at a particular Website, asking every n^{th} visitor to participate in the survey, similar to an election exit poll. Issues of invitation timing result in increased potential for non-response. List-based samples of high-coverage populations start with a frame or list of those with Web access. Email invitations are sent to either to everyone or to a group on the list. This reduces non-coverage issues, but does nothing for non-response rates. In mixed-mode designs with choices of completion method, the Web-based survey is just one alternative for response. This approach is more costly and raises issues of equivalence of measurement across instruments. With pre-recruited Internet user panels, panel members do not self-select, but are recruited using probability sampling methods such as RDD (random digital dialing). Here, non-response can occur at many stages of the recruitment and survey process. The last type of Web-based survey sampling method, probability samples of full populations, provides subjects with the equipment and

tools necessary to participate. Couper [2000] believes it is the only approach that allows generalization beyond the current populations of Internet users. Recruitment response rates are low, but once participating, response rates are quite high.

Given Couper's [2000] topology, coverage error (the mismatch between the general population and the sampling frame) and random sampling within the sample frame are the biggest threats to inference from Web-based surveys to general populations [Couper, 2000]. First, people who participate in online surveys are different than the general population. Second, on the Web, a sampling frame of all online users cannot be identified. Other researchers support Couper's conclusions. For example, Georgia Tech University's Graphic, Visualization, and Usability Center (GVU) found it impossible to draw a random sample from a complete or nearly complete list of Internet users [Kehoe & Pitkow, 1996] which, in addition, makes it impossible to track non-response rates [Kehoe & Pitkow, 1997]. Indeed, some researchers claim that the only way of reliably sampling is through a national census [Robson, 1993], but this method is not suitable for all types of surveys. Also it adds extra cost and access to census information is not readily available.

Additionally, estimates of the online population and its demographics are not consistent among measuring entities [Couper, 2000]. Some studies show that those who participate in Web-based surveys may be more experienced, more intense Internet users, and have stronger Internet skill sets than those who do not participate in the survey [Kehoe & Pitkow, 1997]. They may be predominately male, younger and from households with fairly high incomes [Sheehan & Hoy, 1999; Sohn, 2001]; or consist of more whites, less African Americans and Hispanics than are in the general population [Witte, et al., 2000]. However, recent surveys geared towards examining the demography of the digital divide show that the gap between the number of men and women users has disappeared (NUA, 2001). Economics, age and ethnicity as percentages of the total population continue to have significant gaps although all ages and ethnic groups are now online (NUA, 2000). For example, Yun & Trumbo [2000] found that those who return electronic surveys tended to have high connectedness with their profession, more education, greater number of contacts with other colleagues, a greater volume of email use and more task-related email then those who completed the paper survey. Zhang [2000] found that Web-based survey respondents had a higher self-perceived ability to use the Internet, used the Web more often, were seven years younger in mean age, but did not differ significantly in years of Internet experience, Web access or gender from those who responded using fax or postal mail. In summary, whether or not the Web access will ever be universal remains speculation [Couper, 2000].

The online population is not reflective of the offline population distribution, and it is changing continually. To infer for a general population based on a sample drawn from an online population is not as yet possible and will not be possible until the online and offline populations reflect each other. In the meantime, it is suggested that marketing and promotion techniques may mitigate demographic differences, at least in some characteristics other than gender [Kehoe & Pitkow, 1996].

The inability to identify all online users finds its source in the lack of Internet central registries and the fact that many people have multiple online email addresses. This makes email and Web-based surveys limited in their ability to provide generalizable results due to self-selection, non-random and non-probabilistic sampling [Yun & Trumbo, 2000; Dillman, 2000; Schaefer & Dillman, 1998; Swoboda, et al., 1997; Kehoe & Pitkow, 1997]. For example, one study of U.S. online users chose every 6th name from a list of 55 Internet Service Providers (ISPs). Then the email addresses associated with the selected ISP were generated from Yahoo's email catalog directory service [Sheehan & Hoy, 1999]. This directory contains over 15 million names from three different sources. Email address selection was limited to the Yahoo domain name. During sample selection, sometimes the catalog displayed all names and in other cases only the first "X" number of names. Another attempt at generalization occurred when researchers collected Internet usage hours data in the survey to create subgroups from which people were randomly selected to make statistically valid statements about the Internet population as a whole [Kehoe & Pitkow, 1996].

Online sampling attempts also have applied offline sampling adjustment techniques such as over sampling (sometimes named double or convenience sampling). Over sampling is based on the theory that an overly large sample size may reduce the chances of systematically excluding segments of the population [Kehoe & Pitkow, 1996]. Smith [1997] supports this approach because he found that in email surveys, invalid emails affected sample selection validity and over sampling was used to adjust for the problem. However, over sampling alone is insufficient. Probabilities for selection can be estimated by comparing the sample (post survey) to benchmarks, such as official government statistics by matching subgroups analysis or random sampling techniques (e.g., random digit dialing) with similar demographic data [Kehoe & Pitkow, 1996; Witte et al., 2000]. Witte et al. suggest sensitivity analysis to estimate and determine weighting adjustments for representativeness across subgroups. However, as stated earlier, an online survey sample can not be as yet representative of a general population because online, some demographic groups are strongly over represented and others are underrepresented in the sample that is drawn only online [Witte et al., 2000].

Table 3: Sampling and Subject Selection for Online Research
• Online results generalizable to offline population is unattainable
• Identification of online populations continue to be impractical
• Alternative: Series of indicative, not generalizable, results through artificially defined sampling frames

In summary (Table 3), research results that are generalizable to offline or online populations will continue to be unattainable given the nature of the Internet and the disparities between online and offline populations. The continuing expansion and change of servers that make up the Internet will increasingly make it difficult for researchers to find lists of servers and assess how up-to-date that information is. ISP access policies, email filtering software, multiple email addresses for individuals and increasing volumes of email may cause a decrease in unsolicited email response rates even if it is from legitimate researchers [Sheehan & Hoy, 1999].

The alternative is not to build knowledge through generalization in Internet research, but rather to build knowledge based upon a series of studies that provide indicative data. In this approach, random sampling is limited to a sampling frame that is artificially defined as a set of specific target groups, individuals, discussion groups, etc., then to follow statistical sound random sampling techniques for the artificially defined sampling frame [Coomber, 1997; Yun & Trumbo, 2000].

Distribution Methods and Response Rate Management Guidelines

Obtaining significant response rates with conventional postal surveys has always been a challenge. This situation has not changed for either email or Web-based surveys. As the volume of electronic surveys has increased, the types of populations being studied has become larger, less cohesive and less interested in the technology as the amount of unsolicited non-survey email has increased [Sheehan, 2001, Cho & LaRose, 1999]. Bosnjak & Tuten [2001] identified categories of response types that include: complete responders, unit responders (do not participate at all), answering drop-outs, lurkers (view but do not answer questions), lurking drop outs (view some but not all of the survey), item non-responders (only answer some of the questions, but complete the survey), and item non-responding drop-outs (answer some questions, but drop out before completing).

In many cases, email and Web-based surveys fail to reach response rate levels of postal surveys and may threaten the use of electronic surveys [Couper et al., 1999; Schafer & Dillman, 1998]. Email response rates of 20% or lower are not uncommon [Witmer, et al in Jones, 1999] and, although rates exceeding 70% have been recorded, they are attributed to respondent cohesiveness (e.g. an existing workgroup) as often occurs in organizational studies

[Walsh et al., 1992]. In addition, response characteristics differ between postal, e-mail and Websites. For example, email response is faster then postal survey responses [Sheehan and McMillian, 1999] without significant impacts on survey results [Yun & Trumbo, 2000; Sheehan & Hoy, 1999]. Yun & Trumbo [2000] recommend all three distribution methods, but conclude that if only one can be used and the population is an online population, the Web-base survey supported with various forms of pre-notification is advisable.

There is very little researchers can do to persuade someone to participate if he/she simply prefers not to participate. Incentives and techniques used in postal surveys to increase response rates may not be possible in electronic formats [Couper, 2000]. Technical difficulties alone may keep response rates low [Couper, 2000]. A lack of survey salience (the association of importance and/or timeliness with a specific topic to a potential survey subject) can also reduce responses [Sheehan & McMillian, 1999; Watt, 1999]. Response rates may also be affected by some systematic judgement by a segment of the population being studied, causing them to be excluded from the result [Kehoe & Pitkow, 1996; Sheehan, 2001]. For example, invitations to participate posted on discussion groups may get higher response rates from technical discussion groups because they are more interested in any type of online interaction while those groups dedicated to health support issues may interpret the survey participation request as an intrusion on their privacy. The attrition rate – the number of people who started to take the survey, but did not complete it -- can be used to reveal some systematic judgement by a group [Kehoe & Pitkow, 1996]. For example, attrition rates may be calculated if the survey captures the "link-to" source of the survey. Counts by that source for completed and partially completed surveys provide the basis for rate calculations.

Another reason for high non-response rates may revolve around issues of privacy and confidentiality [Couper, 2000]. Research also suggests that the inability to inspect the survey document prior to completing it, as can be done with a postal survey or with a Web-based survey where all the questions are on one scrolling screen, increases non-response rates [Crawford et al., 2001].

However, survey design and distribution features also affect response/non-response and attrition rates. Experiments comparing short and long email questionnaires did not show that shorter questionnaires produce significantly higher response rates than longer versions [Witmer, et al., in Jones, 1999; Sheehan, 2001]. Cash incentives [Kehoe & Pitkow, 1997, Cho & LaRose, 1999] and chances to win prizes in a lottery [Brick, et al., 1999] have been shown to increase the number of responses twice as much as altruistic motives [Tuten, Bosnjak & Bandilla, 2000]. However, such incentives may introduce a systematic bias into the study. In multi-year repeating

survey studies, incentives were shown to increase the number of completed questionnaires, but not the total number of respondents from year to year [Kehoe & Pitkow, 1997].

Another design feature that seems to affect attrition rates is the location of the request for personal (demographic) data in the survey. Attrition rates were significantly lower when personal data was requested at the beginning of Web-based survey rather than at the end of the survey [Frick et al., 1999]. Placing the data request at the end of the survey presents a surprise to the respondent to which he/she reacts negatively by dropping the survey before completing it. Placing the data request at the beginning may be perceived as honesty on the part of the researcher. This helps to create an atmosphere of greater trust and to build a quality relationship.

How survey subjects are invited to participate in the survey, and how survey completion is encouraged through reminders, can affect response rates. The perceptions of burden (the effort required to complete the survey) can be manipulated and affect non-response and attrition rates [Crawford et al, 2001]. For example, those who were told the survey would take less time, those received an automated (embedded) password, and those who received more frequent reminders, were all more likely to accept the invitation. However, these factors did not have significant effects on signing up for the survey [Crawford et al., 2001]. The use of automatically generated passwords in email invitations to protect against "ballot stuffing" and allow subjects to break off and re-enter to complete a survey can affect attrition and non-response rates. When subjects were given passwords having no ambiguous characters (e.g., 1 [one] or l [el]) their response rates were significantly higher than those subjects with password ambiguities [Couper et al., 2001].

The presentation of the survey through single email containing both a "cover letter" and the survey is likely to cause a strong negative reaction [Witmer et al, 1999; Mehta & Sivadas, 1995; Sheehan, 2001; Cho & LaRose, 1999]. A multi-step process that separates the invitation and survey presentation is recommended. For email surveys, this process includes a short email that pre-notifies and introduces the coming survey. The email requests a reply email if the user declines to participate in the survey ("opt out") or requests a reply if the user wishes ("opt in") to receive the survey [Witmer et al, 1999; Sheehan & Hoy, 1999; Sheehan, 2001]. Follow-up reminder emails after the first publication also appear to spike participation [Smith, 1997; Sheehan & Hoy, 1999]. More sophisticated approaches integrate online and offline contacts. In one study, contact with subjects occurred up to four times, beginning with an invitation postal letter, then a paper survey and an email survey with URL for Web version. These were followed with reminder postcards. Surveys that arrived just before the emails were found to boost

response rates to above 70% [Yun & Trumbo, 2000]. In another study, varying the interval time periods between reminders does not appear to make a difference in response rates [Claycomb et al, 2000].

It appears that Web-based surveys have an advantage over email surveys. When a Web-based survey is preceded by an email inviting individuals to the URL to participate, the Web-based survey outperforms email survey participation significantly [Smith, 1997]. However, there are still unresolved issues. Techniques to increase the potential for subjects to open and read the email invitation must be developed, given the general increase in unsolicited email [Sheehan, 2001]. The language in the subject line, the email address of the sender and the sender's name may all influence whether or not the invitation to participate is read. As with email surveys, multiple contact methods are be useful for Web-based surveys as well. Respondents in a palliative care survey were invited to participate on a Web site, listserv and newsletter. Results showed that 83% of the respondents search the Internet for clinical information, 83% use email, 69% access online medical journals and 59% were subscribers to a palliative-care related listserve or newsgroup [Pereira et al, 2001]. In addition, as discussed earlier, different groups may have dramatically different response rates based on survey salience.

In summary (Table 4), achieving high response rates to an electronic survey depends very much upon how people are asked to participate. Web-based surveys are slightly more flexible than email surveys in that they can be separated from the invitation to participate (e.g., email, advertisement or posting) yet connected to it through the URL link in the invitation. Web-based surveys also allow question grouping and modularization. Partially completed survey data can be saved even if the survey taker abandons the survey before all questions are answered so attrition rates can be calculated. However, the distribution of any electronic survey must be carefully planned.

Table 4: Response Rate Evaluation

Response Rate Impact Items	Email	Web-based
• Technical breakdowns likely	No	Yes
• Survey can be made salient to respondent interests	Yes	Yes
• Systematic judgement by segment of survey population can be prevented	No	No
• Privacy and confidentiality issues likely	Yes	Yes
• Possible to inspect entire survey before completion	Yes	Probably not
• Possible to offer financial incentives	Not easily	Not easily
• Requests for personal data first	Yes	Yes
• No use of ambiguous characters in passwords	Not needed	Yes
• Multi-step invitation and survey presentation process	Yes	Yes
• Periodic reminders to complete the survey	Yes	Yes
• Appropriate subject line in email invitation	Yes	Yes
• Multiple ways to contact & invite respondents	Yes	Yes
• Customization to target population – invitation language, type of notification media, and follow-up process	Yes	Yes
• Save responses to partially completed survey	No	Yes

The language used in the invitation, the type of notification media and the follow-up process (e.g., the type, and number, of reminders) must be customized to the target population. Distribution techniques can and will invade privacy of some individuals and groups online. Therefore, no detail should be overlooked.

Survey Piloting Guidelines

Motivating subjects to complete a survey increases as question difficulty increases (e.g., question interpretation, data entry volume, number of choices), respondent's ability to answer decreases (e.g., perform complex mental tasks, make judgements), and respondent's motivation decreases (e.g., topic salience, belief in usefulness of questionnaire) [Krosnick, 1999]. Therefore, survey piloting is crucial to achieving research goals and ensuring that subjects complete the survey. To quote a leader in survey development, "Survey piloting is the process of conceptualizing and re-conceptualizing the key aims of the study and making preparations for the fieldwork and analysis so that not too much will go wrong and nothing will have been left out" [Oppenheim, 1992, p. 64].

As stated earlier, going online with a survey does not preclude poor design. The design of response alternatives (open and closed questions, frequency scales, reference periods, and rating scales) and question context (researcher's epistemic interest, and adjacent questions) can create bias that may destroy the quality of any survey [Schwarz, 1999; Krosnick, 1999]. Inattentiveness to detail also inhibits quality. For example, Preece et al. [2002] found common errors in electronic surveys to include (1) requests for exact demographic data when a range would be more appropriate (e.g., age), (2) overlapping scales (e.g., 1-3, 3-6), (3) missing and inaccurate instructions, (4) the use of specialist terms, and (5) insufficient space for open ended question answers. Other errors are subtler. For example, inaccurate motivational techniques embedded in the survey or the invitation (e.g., estimated completion time, progress indicators calibrated to show time to completion) may create distrust and subsequently increase abandonment [Crawford et al., 2001]. The way a question is contextually framed and worded may encourage subject acquiescence in responses (e.g., attempting to anticipate correct answers or answers expected by the researcher) [Krosnick].

In discussing the piloting of paper based surveys, Oppenheim [1992] stresses that piloting begins with question development because every question, every question sequence and every scale used in the survey must be tested and re-tested as improvements are made. In addition to the questions, the question lay-out, instructions to

respondents, the answer categories and even the question numbering systems should be tested along with the sampling and data analysis techniques.

Researchers [Dillman, 2000; Schwarz & Sudman, 1996] have developed numerous procedures for survey pretesting. Dillman [2000] suggests a multi-stage testing process that integrates testing techniques and can be applied to either paper or electronic surveys (Table 5). The process begins after the survey is considered "ready" by its developers. Stage 1 consists of a review by knowledgeable colleagues and analysts to ensure question completeness, efficiency, relevancy, and format appropriateness. In Stage 2 cognitive pretesting consists of observation and "think aloud" protocols while a respondent completes the survey and is followed with a retrospective interview. This evaluates cognitive and motivational qualities of the survey. This helps to ensure wording understandability, interpretation consistency, logical sequencing, and overall positive impression from the look and feel of the survey. Stage 3 consists of a small pilot study that emulates all the procedures proposed by the main study. Dillman [2000] suggests, for large surveys, that a sample of 100-200 individuals complete the survey and that the resulting data be analyzed to determine opportunities and needs for question scaling improvement, to reduce the number of questions because of high correlation, to eliminate or change questions with high non-response rates, to test if open ended questions provide useful information, and to estimate response rates. In the last stage, Stage 4, researchers conduct one last check using people who have no connection to the survey. The objective is to catch typos and errors that may have been inadvertently introduced during the last revision process. This approach reflects similar pilot testing approaches used in systems and Web-development [Preece et al., 1994] as well as conventional survey development [Oppenheim, 1992].

Table 5: A Survey Pilot Process

Stage One:	***Stage Two:***	***Stage Three:***	***Stage Four:***
Survey by knowledgeable colleagues to ensure question completeness, efficiency, relevancy and format appropriateness	Observation and "think aloud" protocols test respondents complete survey. This is followed with retrospective interviews.	Small pilot study that emulates all the procedures proposed by the main study.	Last check by non-researchers for typos and errors inadvertently introduced during the last revision process

Through online survey piloting, researchers have encountered unanticipated consequences that either delayed or aborted their studies [Smith, 1997; Witmar et al., 1996]. Witte et al. [2000] found that their survey was too long. As a result, they modularized it offering only one module (randomly) to subjects with the option of

completing additional modules at the end of the survey. Seventy percent of their subjects completed the additional modules. Web specific question structuring problems have been revealed through piloting [Preece, 2002]. These problems include but are not limited to such items as: too many open-ended questions, incorrect defaults (hidden or revealed), large enough text boxes that scroll, question independence (so one mistake does not invalidate the complete survey), ambiguous wording, inconsistent terminology, non-orthogonal categories, overlapping categories, and answers that can't be undone.

Piloting can also reveal undeliverable email, declined, and completed survey rates which are useful for estimating the amount of over sampling that may be required to attain a valid sample size for a particular confidence level [Sheehan & Hoy, 1999]. Web technology allows researchers to track and analyze survey non-responses by question using log files to understand why a survey is not being completed [Bosnjak & Tuten, 2001]. This may be helpful in testing question sensitivity, clarity and understandability. To use this technique, each question must be displayed separately on a screen, participants are not forced to provide an answer before moving on, and each page of the questionnaire must be downloaded separately from the server and not allowed to reside in the Web browser's cache memory. This makes for an awkward and time-consuming survey taking process. In one pilot case, subjects consistently abandoned the survey at the same point allowing researchers to identify unclear instructions as the reason for survey abandonment [McCoy & Marks, 2001]. In the second case, the abandonment rate was 5% and no major flaws were revealed in instructions or format [McCoy & Marks, 2001].

Table 6: Frequent Mistakes Caught through Piloting

- Bias in question/answer wording
- Inconsistent wording and spelling errors.
- Requesting inappropriate demographic data
- Overlapping question scales or selection options
- Inaccurate or missing instructions
- Technical vocabulary with no definitions
- Insufficient space for open-ended question answers
- Lack of motivational techniques to go to the survey and/or complete it

In summary (Table6), a conscientious and complete pilot of the survey and the survey distribution and data collection process can help to avoid painful mistakes that can ruin an important research project.

Guideline Summary

It appears that a Web-based survey is the most appropriate format for online data collection when research costs are a constraint, timeliness is important and the nature of the research requires it. However, this method does

present both technical and administration challenges that do not exist with postal or email surveys. Past research provides many insights for designing, testing and using Web technologies. However, sampling and distribution remain challenges to any online survey research effort. Careful analysis of the study population should be conducted to determine whether probability-based or non-probability-based approaches to sampling are appropriate and whether mixed media should be used to reach offline populations as well as online populations should be sampled using other survey formats. In all cases, full recognition of the limitations of the sampling method must be well documented. Piloting is a must to perfect the survey as much as possible because response rates will be low. A participant reminder plan is essential in addition to a carefully constructed invitation to participate that motivates the target population to participate. Confidentially and privacy must be assured. Subjects should be informed regarding the study itself, how the results will be used, and reassured regarding the credibility of the researchers. A third party guarantee should be obtained if possible.

Because online surveys are so different from postal surveys, many issues regarding Web-based online surveys remain unresolved, providing ample opportunities for continuing research. For survey practitioners, current survey knowledge can be used with caution as a preliminary set of guidelines for survey development. The more we use Web-based surveys systematically in research and report those findings, the more gaps in our knowledge of this field will be filled.

Guidelines Application: Research Background

The current research explores and defines the nature of participation in online discussion groups, especially those aspects of "non-public participation" commonly referred to as "lurking." When lurking is defined as "not posting to a discussion group," the mean level of "non-public participation" for all discussion groups is lower than the previously reported 90% and the volume of lurking can vary dramatically among different online discussion groups [Nonnecke & Preece, 2001]. For example, health-support discussion groups have, on average, significantly fewer lurkers (46%) than software support discussion groups (82%) [Nonnecke & Preece, 2000].

As recommended in Sudweeks & Simoff [1999], this research sought to balance the strengths of quantitative and qualitative methods to provide multiple perspectives that can be refined and integrated into a single model. Preceding the quantitative logging study [Nonnecke & Preece, 2001] a qualitative study was conducted to collect preliminary data as to why peripheral members of online communities "lurk." It consisted of ten face-to-face

semi-structured interviews. The results guided discussion group selection for the logging study as well as the current study to examine causal factors. The interview results indicated that non-public participation is a strategic activity. Researchers were able to propose a preliminary model, called the gratification model, to categorize the reasons for lurking. It suggests that support and information gathering needs can be met through non-public participation [Nonnecke & Preece, 2001]. Seventy-nine stated reasons for non-public participation were grouped into four categories: member characteristics, group characteristics, membership life cycle stage, and external constraints [Nonnecke & Preece, 2001]. This study, a quantitative online survey, will be used refine and validate the proposed model of the qualitative study. The research is seeking answers to the following questions:

1. Do age, gender, education level, experience with Internet, experience with online community technologies, work status, and/or work environment affect the amount of lurking one does and the reasons for doing it?
2. How does membership in multiple online communities, frequency of access, the type of community or the technology used by the community influence the amount of lurking one does or the reasons for it?
3. What, if any, are the relationships between lurking and stage of membership within an online community (e.g., joining and leaving an online community)?
4. What is the relationship of "feeling like a member of a community" and lurking?

The researchers were faced with many decisions regarding design and distribution of the survey. As researchers made decisions regarding survey design, subject privacy and confidentiality, sampling and subject selection, distribution and response rate management, and survey piloting, they used the knowledge from the literature (a complete reference list is included). In addition, these assumptions about the nature of the behavior under study and the scope of their research were made:

- All potential survey subjects are Internet users
- Getting people who do not post in discussion groups to respond to a survey may be problematic
- Subject identities are not required for anything other than follow-up interviews of a small subset of respondents to clarify response pattern clarification, if needed
- Topic salience was going to be a major problem for almost all respondents (i.e. reasons for joining or leaving a discussion group are, most likely, of no interest to members of a sports discussion group, for example)
- Using the population of all online discussion groups as the universe from which a representative random sample is drawn is not possible

- Researchers' university affiliations and reputations must be leveraged to establish survey credibility
- Funding to incent survey participation is not available
- The survey could easily become lengthy based upon the previous qualitative study results
- Results from the qualitative survey will guide the language used in the online survey

After reviewing the literature on electronic surveys, their first decision was to use a web-based survey using email invitations and reminders posted in public online discussion groups. The research team had a working knowledge of survey development, web and database technologies as well as access to additional technical resources.

Guidelines Application: Survey Design Decisions

Cold Fusion, Microsoft Access and HTML were selected for the technical design. This allows researchers to (1) support multiple platforms and browsers [Yun & Trumbo, 2000], (2) prevent multiple submissions [Kehoe & Pitkow, 1996], (3) provide multiple opportunities for saving respondent answers [Smith. 1997], (4) collect both coded and open-ended responses [Yun & Trumbo, 2000], and, (5) provide immediate "thank-you" feedback upon survey completion [Smith, 1997]. This technical approach also provides the ability to track respondent identity for follow-up interviews if they "opt in" to follow-up interviews and, at the same time, protect individual privacy without the use of cookies [Cho & LaRose, 1999]. Also, if the survey proves, in piloting, to require more than 10 minutes to complete, it allows researchers the option of providing re-entry access using non-cookie passwords. The technology also allows a respondent who belongs to more than one community in the study to complete the survey for each community while at the same time prevent "ballot stuffing." The processing logic produces a gently worded error message: "It appears you have already completed a survey for this online community. Please contact the survey administrator at the link below to investigate the problem", if ballot stuffing occurs.

The survey is designed to have a professional, simple layout using a straightforward navigation strategy, keeping graphics and color to a minimum in an effort to add credibility to the survey as well as keep downloading time as short as possible [Couper et al., 2001; Dillman et al., 1998; Preece et al., 2002]. The survey, in its current state of design, has 28 primary questions, about 20 sub-set questions, and 12 demographic items. Researchers, following an introduction page, divided these questions into three sections, each having a "submit" and "save" function: 1) demographic questions, 2) questions related to the discussion group where invitation to participate was posted, and, 3) questions related to a discussion group that was permanently left. If the respondent abandons the

survey, the data from the completed sections is not lost. This was a compromise between having the whole survey on a single page vs. displaying each question on its own page. The download and submit processing time required for over 50 single pages or one very large page was considered too burdensome. When a survey is completed, the respondent immediately thanked and notified that his/her survey was successfully completed.

The survey begins with a single introduction page. Like the email invitation that brought the respondent to the survey website, its purpose is to establish a trusting relationship with the prospective respondent that encourages him/her to proceed into the survey. To do this, there is text to (1) establish the authority and credibility of the researchers, (2) explain the survey purpose, (3) explain benefits of the results to online communities to address the salience issues of the survey, (4) establish respondent confidentiality and privacy, (5) provide open access to researchers through email address links to answer questions before starting the survey, (6) explain the sampling methodology, and, (7) provide a third party guarantee of the survey's authenticity and credibility using the Institutional Review Board approval with supporting links [Cho & LaRose, 1999].

Before demographic information is collected, a page for "opt-in" informed consent is presented along with links to small pop-up windows to display term definitions. For example, the terms 'active', 'occasional', 'join', 'participate', 'leave', 'member' and 'visitor' have links throughout the survey. These links exist in the survey questions also. A small incentive, the survey results will be available to respondents, to participate in the survey is provided on this page. By providing an email address, researchers will be able to notify the respondent of results availability. Respondents are also asked to "opt-in" to a follow-up telephone interview. Researchers explain that they intend to randomly sample a few respondents to explore results patterns more deeply, if needed.

Researchers are comfortable with asking both coded and open-ended questions, but they are limiting the open-ended questions to optional opportunities to add information at the end of a coded question set and are using text-input boxes with wrapping and scrolling, not single line entry [Preece et al., 200l; Stanton, 1998; Andrews et al., 2001]. Skipping these questions does not affect the coded survey results. Coded questions use nominal scales, Likert scales, semantic differential scales, single and multiple choice selection options [Oppenheim, 1992].

Following Dillman's [2000] four stage development process, survey design required four rounds of prototype development before researchers felt they could proceed to stage 2 – cognitive pretesting. Unlike paper surveys where questionnaire presentation is stable, web-based survey question presentation requires the extensive use of HTML tables to control layout, wording and selection option alignment with testing on numerous browsers

and preferences within browsers [Preece et al., 2002]. This was particularly important for any scales where a shift in alignment can cause misinterpretation of the question or make it unanswerable.

Question language proved more challenging than first anticipated by the researchers. In addition to maintaining question objectivity to control for bias, shorter sentences are better for reading on the screen. As Nielsen [2000] and others have demonstrated, people do not read web pages, they scan them, looking for key word and phrases. Therefore, survey questions and instructions became briefer as researchers reviewed the prototyped screens. For example, the original statement *"This second set of questions is similar to the first set, but focus on an online group which you have permanently left and no longer consider yourself to be a member of.*" became "*The questions below pertain to an online group you have permanently left."* There is a constant struggle to maintain the balance between brevity and a friendly tone.

Researchers also had to work to eliminate redundant questions and refine the ones kept. Non-response rates are anticipated to be due to the continual growth of email and use of electronic surveys [Couper et al., 1999; Schafer & Dillman, 1998]. They anticipate a high attrition rate if the survey was too long or irrelevant to the respondent. The online community shared interests (e.g., stock market, dieting, soap operas, health) are not those addressed in the survey. The survey is collecting data about "the different ways people use and participate in online groups." Therefore, each review round resulted in eliminating questions and revising the introduction and invitation.

Another challenge in question development faced by the researchers was how to present the 79 reasons and four categories for being attracted to, participating and/or leaving an online group generated from the initial interview-based study. The original idea was to use the results of the qualitative study directly, but researchers were concerned that the categories could be limiting and prototype testing proved that 79 items are overwhelming to review and select from on a single screen display without constant scrolling up and down. At the end, the researchers removed all items that could be considered duplications, put similar items together into groups of 3 and 4, and removed all category headings. An open-ended question, "Please report any other reasons you might have…" was added. When all of this was completed, the survey was considered ready for stage 2 -- cognitive pretesting.

Guidelines Application: Subject Privacy and Confidentiality Decisions

To protect privacy and reduce intrusion, researchers decided to post the survey invitation only to public online discussion groups rather than email individual members. Direct emailing to each group member would have

provided a mechanism for tracking individual responses, but was considered overly intrusive and for some participants would have been considered spam [Cho & LaRose, 1999]. In addition, obtaining email addresses for each group member is becoming increasingly difficult. Respondent identity will be obtained only if he/she opts to provide it. The original intention was to post the invitation only after obtaining permission from the discussion group owner, but this proved problematic in the pilot work. The invitation will be posted without owner permission, unless discussion group policies directly require owner permission to post a message that is not directly "on topic." The invitation text (1) explains the nature of the posting, (2) builds researcher credibility and authority, (3) demonstrates third party guarantee of trustworthiness by mentioning IRB approval, (4) explains discussion group selection methodology, and, (5) explains how taking the survey may benefit the potential respondent [Cho & LaRose, 1999]. After reading the invitation, discussion group readers can ignore the post or self-select to take the survey when they click on the survey URL in the invitation. The IRB process ensures that subjects understand what they are participating in, are told of any known risks, and requires documented subject acceptance to participate in the research before research is conducted.

An "opt-in" informed consent approach was selected for this survey, as mentioned earlier. If the respondent click on "I do not accept" then proceeds with the study, the data will be permanently removed from the database. Under age respondents are asked to provide an email address of an adult who can give consent. Researchers will follow-up to obtain consent. If consent is not given, the responses of the underage respondent will be removed from the database before analysis begins.

To provide additional protection level of respondent privacy, participation in the follow-up sample is completely optional through "opt-in" selection. In addition, an email address is not required and cookies are not used. As a result, survey tracking is at the discussion group, not the individual level. If an email address is provided, this identity data is stripped from the main database before analysis begins and is accessible only by the research team in a separate database table.

Guidelines Application: Sampling and Subject Selection Decisions

Acknowledging the continuing decrease in response rates as online surveys proliferate, researchers anticipate low response rates from the target population. For this reason an easily replicated population definition and sampling approach was developed [Couper, 2000; Coomber, 1997; Yun & Trumbo, 2000]. Researchers decided

to use the probabilistic sampling method that begins with knowledge of the target population to permit the measurement of non-response at the discussion group level. However, the target population is not the total population of discussion groups, which is impossible to identify. Therefore, the results from this research will be considered indicative and no attempt will be made to infer to the general population of discussion group readers. Researchers hypothesize, based upon the high degree of variation in lurking among types of discussion groups, that there may be great variation in responses among readers in different discussion groups [Nonnecke & Preece, 2000]. Given the decision that results will only be indicative, not predictive, researchers created a sampling process that can be replicated across numerous discussion group populations. This will provide multiple, comparable indicative results which may be all Internet researchers can hope to attain given the ubiquitous and changing nature of the Internet. The following process documents the sampling process.

Step 1: Select a population to be sampled

This study is limited to asynchronous discussion groups because the previous studies were based asynchronous discussion groups. Discussion groups are aggregated by many different portal and non-portal resources such as MSN, Yahoo!, Catalist, Talkcity, Google, Altavista, http://webcom.com/impulse/list.html, and http://tile.net/lists as well as many websites that aggregate discussion group of a particular character or interest. Because previous research clearly demonstrated that different types of discussion groups have different lurker characteristics, an aggregation of heterogeneous discussion groups was determined to be ideal to continue the study of diversity. The "MSN web communities" have such diversity and was selected for that reason. There are 16 discussion group categories at the highest level of the MSN's community hierarchy. Twenty-five percent (25%) of these categories were selected using a random number generator to narrow the sampling frame. The categories selected were (1) health and wellness, (2) government, (3) sports & recreation, and (4) organizations. This population was further narrowed to ensure that discussion groups had sufficient critical mass (at least 50 members), were open to public participation, and were not just mailing lists, but active discussion groups (4-5 people posted within the past 90 days). A total of 1304 discussion groups were identified as members of this target population.

Step 2: Select a stratified random sample

Given this population, it is now possible to select a random sample from the frame. A stratified random sampling approach was used to ensure that

Table 7: Stratified Sample within the Sampling Frame

Category	Groups Meeting Criteria	Pop. %
Health & Wellness	435	33%
Government	139	11%
Sports & Recreation	531	41%
Organizations	199	15%
Total	1304	100%

each category was proportionally sampled (Table 1). If a category had additional hierarchical clustering of discussion groups, these were ignored for sampling purposes. All groups meeting the criteria were counted within that group as if there was no categorization below the highest level.

If this survey research is replicated with a non-categorized, but a heterogeneous discussion group population, then an inspection method should be added to categorize the discussion groups to ensure proportional representation in the sample. To attain a 95% confidence level that the sample results are inferable to the sampling frame, 359 discussion groups need to be surveyed. However, knowing the response rate may be less than 50%, over sampling was estimated to compensate for this lack of coverage. For this reason, the sample was adjusted to 371 discussion groups.

To measure non-response rates, the survey captures the name of the discussion group as entered by the respondent and/or the source URL of the discussion group on which the invitation was placed. This will also allow results to be analyzed by category. If one category appears to be proportionally underrepresented in the response rates, additional samples can be pulled in for the under-represented category. The process can be repeated using various discussion group aggregations with this sampling method until researchers decide that a sufficient number of survey responses have been collected for analysis.

Guidelines Application: Distribution Methods and Response Rate Management Decisions

Based on the way the survey is designed and the technology is applied, researchers will be able to estimate the non-response rate at the discussion group level and will be able to calculate the attrition rates at the respondent level by survey section.

The invitation to participate in the research is separate from the survey itself [Witmer et al., 1999; Mehta & Sivadas, 1995; Sheedan, 2001; Cho & LaRose, 1999]. It has been designed to build a trusting relationship from the beginning of the survey experience. A customized invitation will be posted to each group in the sample. Knowing that the majority of responses to electronic invitations occur very shortly after invitation posting, reminders will be posted each week for three weeks following the initial invitation [Yun & Trumb, 2000; Claycomb et al., 2000]. By that time the invitation should have been read by all readers who visit the discussion group at least once a month. The database will be examined before each reminder to determine if a reminder is warranted. If, after three

reminders, few responses from a particular discussion group have been recorded, a decision may be made to select another discussion group for the sample.

Many design features are used to reduce attrition. The survey introduction will use a full disclosure, direct access to researchers, and third party guarantor (IRB) to built trust and credibility in the researchers. Demographic data was gathered at the beginning of the survey [Frick et al., 1999]. A realistic estimate of the time required to complete the survey, a description of the survey structure and indicators of survey progress (using static statements) are provided.

Guidelines Applications: Survey Piloting Decisions

Following Dillman's [2000] four stage piloting process, Stage 1 – initial survey development is complete, as is Stage 2 – cognitive pretesting although a slightly different approach was taken that is discussed by Dillman [2000] and Preece et al [2002]. The researchers constructed draft survey questions using a word processor. One researcher developed the online prototype, which went through two rounds of review with colleagues to ensure question completeness, efficiency, relevancy and format completeness. Stage 2 consisted of several subjects, not involved in the research, who completed the survey under the observation of a researcher using "think out loud" protocols with retrospective interviews. These cognitive pretests resulted in language simplification on the invitation and survey questions, changes in sequencing, and feedback on the look and feel of the survey. After the prototype was updated once more, an invitation to review the survey was placed on the AoIR listserve. Over 50 people completed the survey and 15 people provided email feedback to varying degrees of detail. This pretesting produced an array of technical testing changes to privacy and confidentiality language and requirements, numerous recommendations for question wording, inconsistencies among questions and elimination of several questions.

The survey and the invitation will soon undergo final changes so that a Stage 3 pilot test of the sampling technique with "live" discussion groups can be conducted. The language must encourage non-public participants in discussion groups to participate in the survey without alienating more active discussion group members. The invitation will be posted on approximately 30 discussion groups not included in the sample but included within the MSN online communities' portal. This will allow researchers to monitor for negative reactions to the posting, estimate non-response and attrition rates and test analysis procedures before going to Stage 4 – last format review before the full study is conducted.

Conclusions: Trade-off Decisions are not Eliminated with Survey Guidelines

All the researchers involved in the developed of this web-based survey are seasoned academics and professionals in their representative fields. Yet, despite this, they are continually challenged by the trade-off decision making required at every step of the survey process. Among their chief concerns are sampling frame and sample selections and building a trusting relationship with prospective respondents. Every review uncovers new apparent weaknesses that require still more adjustments in either the survey or the distribution method. It is hoped that from this process, the researchers can not only address the issues of the subject of their research (i.e., further knowledge of lurkers and lurking), but also add to the knowledge about building and using web-based surveys.

References:

_____. Papers and WWW user surveys methodology since 1994: http://www.cc.gatech.edu/gvu/user_surveys/

Andrews, D., Preece, J., & Turoff, M. (January 2001). "A Conceptual Framework for Demographic Groups Resistant to Online Community Interaction." Presented at the HICSS 2001 Conference. Honolulu, HI.

comscore. (February, 2001). "Online consumers proving erratic." NUA. Available at: http://www.nua.com/surveys/index.cgi?f=VS&art_id=905356472&rel=true

Bachmann, D., & Elfrink, J. (1996). Tracking the progress of email versus snail-mail. *Marketing Research* 8(2), 31-35.

Birnbaum, M.H., (May, 2000) "SurveyWiz and FactorWiz: JavaScript Web pages that make HTML forms for research on the Internet." *Behavior research methods, instruments and computers,* 32(2), 339-346.

Burgoon, J.K., Parrott, R., LePoire, B.A., Kelly, D.L., Walther, J.B., & Perry, D. (1989). "Maintaining and restoring privacy through communication in different types of relationships. *Journal of Social and Personal Relationships* 6, 131-158.

Cho, H. & LaRose, R. (Winter, 1999). "Privacy issues in Internet surveys." *Social Science Computer Review*, 17(4), 421-434.

Coomber, R. (June 30, 1997). "Using the Internet for survey research." *Sociological Research Online*, 2(2), 14-23.

Couper, M.P. (2000). "Web-based surveys: A Review of Issues and Approaches." *Public Opinion Quarterly* 64, 464-494.

Couper, M.P., Traugott, M. W., & Lamias, M.J. (2001). "Web survey design and administration." *Public Opinion Quarterly* 65, 230-253.

Couper, M.P., Blair, J. & Triplett, T. (1998). "A comparison of mail and e-mail for a survey of employees in federal statistical agencies. *Journal of Official Statistics* 15(1), 39-56.

Dillman, D.A. (2000.) *Mail and web-based survey: the tailored design method.* New York: John Wiley & Sons.

Frick, A., Bachtinger, M.T. & Reips, U-D. (1999). "Financial Incentives, personal information and drop-out rates in online studies." In U-D. Reips et al. (eds.). *Current Internet Science. Trends, techniques, results.* Available: http://www.dgof.de/tband99/pdfs/i_p/knapdf.

Kehoe, C.M. & Pitkow, J.E. (1996). "Surveying the territory: GVU's five www user survey's." *The Worldwide Web Journal 1(3).* 77-84. [Also on-line]. Available: http://www.cc.gatech.edu/gvu/user_surveys/papers/w3j.html.

Dillman, D., Totora, R.D., Conradt, J. & Bowker, D. 91998). "Influence of plain versus fancy design on response rates for web-based surveys." Paper presented at annual meeting of the American Statistical Association, Dallas, TX.

Kehoe, C.M., Pitkow, J.E. & Morton, K. (1997). "Eighth WWW user survey [online-line]. Available: http://www.gvu.gatech.edu/user_surveys/survey-1997-10.

Kiesler, S. & Sproull, L.S. (1986). "Response effects in the electronic survey. *Public Opinion Quarterly* 50, 402-413.

Knapp, F., & Heidingsfelder, M. (1999). "Drop out analysis: The effect of research design." In U-D. Reips et al. (eds.). *Current Internet Science. Trends, techniques, results.* Available: http://www.dgof.de/tband99/pdfs/i_p/knapdf.

Krosnick, J. A. (1999). "Survey Research." *Annual Reviews: Psychology* 50(1), 537-567.

Lazar, J. & Preece, J. (1999). Designing and implementing web-based surveys. *Journal of Computer Information Systems* xxxix(4), 63-67.

Loke, S.D., & Gilbert, B.O. (1995) Method of psychological assessment, self disclosure, and experiential differences: A study of computer, questionnaire and interview assessment formats. *Journal of Social Behavior and Personality* 10, pp 255-263.

McCoy, S., & Marks, P.V., Jr. (August, 2001). "Using electronic surveys to collect data: experiences from the field." Paper presented at the AMCIS conference. Boston, MA.

Mehta, R. & Sivada, E. (1995). Comparing response rates and response content in mail vs. electronic mail surveys. *Journal of the Market Research Society,* 17(4), 429-440.

Nielsen, J. (2000). *Designing Web usability.* Indianapolis: New Riders.

Nonnecke, B. (2000). *Lurkers in email-based discussion lists.* South Bank University, London.

Nonnecke, B. & Preece, J. (2000). "Lurker demographics: counting the silent." *Human Factors in Computing Systems*. Paper presented at CHI'00. The Hague, Holland.

Nonnecke, B. & Preece, J. (August, 2001). "Why lurkers lurk." Paper presented at the AMCIS conference. Boston, MA.

Norman, K., Friedman, Z., Norman, K. & Stevenson, R. (2001). "Navigational issues in the design of online self-administered questionnaires." *Behaviour & Information Technology*, 20 (1), 37-45.

NUA. (2001). "Women now outnumber men on the Internet. http://www.nua.ie/surveys/index.cgi?f=VS&art_id=905356873&rel=true.

NUA. (2000). Income & Age Are Largest Gap in Digital Divide. http://www.nua.ie/surveys/index.cgi?f=VS&art_id=905355848&rel=true.

Oppenheim, A.N., (1992). *Questionnaire Design, Interviewing and Attitude Measurement*, New York: Pinter Publishers.

Paolo, A.M., Bonamino, G.A., Gibson, D., Patridge, T. & Kallail, K. (2000). "Response rate comparisons of email and mail distributed student evaluations." *Teaching and Learning in Medicine,* 12(2), 81-84.

Pereira, J., Bruera, E., & Quan, H. (Spring, 2001). "Palliative care on the net: An online survey of health care professionals." *Journal of Palliative Care* 17(1), 41-45.

Preece, J., Rogers, Y. & Sharp, S. (2002). *Interaction Design: Beyond Human-Computer Interaction*. New York: John Wiley & Sons.

Preece, J. (1994). *Human Computer Interaction.* Reading, MA: Addison-Wesley.

Robson, C. (1993). *Real World Research*. Oxford, UK: Blackwell.

Rogers, E. (1987). "Progress, problems and prospects for network research: Investigating relationships in the age of electronic communication technologies." Clearwater Beach, FL: Sunbelt Social Networks Conference.

Schaefer , D.R. & Dillman, D.A. (1998). "Development of standard email methodology: Results of an experiment." *Public Opinion Quarterly* 62(3), 378-397.

Schillewaert, N., Langerak, F. & Duhamel, T. (1998). Non probability sampling for www surveys: A comparison of methods. *Journal of the market Research Society* 4(40), 307-313.

Schwarz, N. (1999). "Self-reports: How the questions shape the answers." *American Psychologist* 54(2), 93-105.

Schwarz, N. & Sudman, S. (1996). *Answering questions: Methodology for determining cognitive and communicative processes in survey research*. San Francisco: Jossey-Bass.

Sheehan, K. B. & Hoy, M. B.. (March, 1999). "Using e-mail to survey internet users in the united states: ethodology and assessment." *Journal of Computer Mediated Communication.*

Sheehan, K. B. & McMillan, S.J. (1999). "Response variation in email surveys; An exploration. *Journal of Advertising Research.* 39(4), 45-54.

Sheehan, K B. "E-mail survey response rates: a review." (January, 2001). *Journal of Computer Mediated Communication*.

Smith, C. B. (June, 1997). "Casting the net: surveying an Internet population." *Journal of Computer Mediated Communication.*

Stanton, J.M. (1998). Am empirical assessment of data collection using the Internet. *Personnel Psychology* 51(3), 709-726.

Sudweeks, F. & Simoff, S.J.(1999). "Complementary explorative data analysis: the reconciliation of quantitative and qualitative principles." In *Doing Internet Research: Critical Issues and Methods for Examining the Net*. S. Jones (editor). Thousand Oaks, CA: Sage.

Swoboda, S. J., Muehlberger, N., Weitkunat, R. & Schneeweiss, S. (1997). Web-based surveys by direct mailing: An innovative way of collecting data. *Social Science Computer Review* 15(3).

Taylor, H. (Winter, 2000). "Does Internet research work? Comparing online survey results with telephone survey." *International Journal of Market Research*, 42(1), 51-63.

Tuten, T.L., Bosnnjak, M., & Brandilla, W. (2000). "Banner – advertised web-based surveys." *Marketing Research*, 11(4), 17-21.

Walsh, J.P., Kiesler, S., Sproull, L.S., & Hess, B.W. (1992) Self-selected and randomly selected respondents in a computer network survey. *Public Opinion Quarterly,* 546, 241-244.

Watt, J.H. (1999). "Internet systems for evaluation research." In G. Gay & Bennington (eds), *Information technologies in evaluation: social, moral epistemological and practical implications*, 23-44. San Francisco: Josey-Bass, no. 84.

Weible, R. & Wallace, J. (1998). "The impact of the Internet on data collection." *Marketing Research,* 10(3), 19-23.

Witmer, D.F., Colman, R.W. & Katzman, S.L. (1999). "From paper-and-pencil to screen-and-keyboard." In *Doing Internet Research: Critical Issues and Methods for Examining the Net*. S. Jones (editor). Thousand Oaks, CA: Sage.

Witte, J.C., Amoroso, L.M., Howard, P.E.N. (Summer, 2000). "Research methodology—Method and representation in Internet-based survey tools—Mobility, community, and cultural identify in Survey 2000." *Social Science Computer Review*, 18(2), 179-195.

Yun, Gi Woong and Trumbo, Craig W. (Sept. 2000). "Comparative response to a survey executed by post, e-mail, & web form." *Journal of Computer Mediated Communication*.

Zhang, Y. (January, 2000). "Using the Internet for survey research: A case study." *Journal of the American Society for Information Science*, 51(1), 57-68.

Food Quality and Preference 102 (2022) 104676

Contents lists available at ScienceDirect

Food Quality and Preference

journal homepage: www.elsevier.com/locate/foodqual

Factors affecting data quality of online questionnaires: Issues and metrics for sensory and consumer research

Sara R. Jaeger [a,*], Armand V. Cardello [b]

[a] The New Zealand Institute for Plant and Food Research Limited, Mt Albert Research Centre, Private Bag 92169, Auckland 1142, New Zealand
[b] A.V. Cardello Consulting and Editing Services, Framingham, MA, USA

ARTICLE INFO

Keywords:
Data quality measures
Online surveys
Consumer questionnaires

ABSTRACT

Methodology used in sensory and consumer research has changed rapidly over the past several decades, and consumer research is now frequently executed online. This comes with advantages, but also with the potential for poor quality data due to uncontrolled elements of survey administration and execution. Published papers that utilize online data rarely report metrics of data quality or note precautions that were taken to ensure valid data. The aim of this paper is to raise awareness of the factors that influence online data quality. This is achieved by 1) identifying factors that can impact the reliability and validity of consumer data obtained with online questionnaires, 2) highlighting indices of online questionnaire data quality that can be used by researchers to assess the likelihood that their data are of poor quality, and 3) recommending a number of indices, counter-measures and best practices that can be used by sensory and consumer researchers to assess online questionnaire data quality and to ensure the highest degree of reliability and validity of their data. By making researchers aware of these sources of invalidity and by presenting available remedies and good practices, it is our intention that sensory and consumer scientists will adopt these measures and report them as a routine part of publishing online questionnaire data. Specific suggestions are offered regarding the important elements of data quality that should be included in scientific manuscripts.

1. Introduction

1.1. Online questionnaires in sensory and consumer research

Methods used in sensory and consumer research have changed rapidly over the past decades. Previously, most sensory and consumer product research was conducted as central location tests (CLT), home-use tests or in field settings. However, the rapid expansion of the internet, email, and online social platforms, along with the proliferation of personal computers, tablets and smart phones, has led to an increasing number of sensory and consumer tests that are executed online. In addition, the ongoing COVID-19 pandemic, which has required social distancing, masking and other precautions, has made CLTs difficult or impossible to conduct. These developments have led researchers to conduct online testing, using a variety of questionnaire formats and on-line platforms. Many of these questionnaires use convenience samples accessed through email lists, available data bases, or crowdsourcing platforms. Still others employ vendor panels, i.e., consumers who have registered to serve as consumer respondents for surveys administered by a commercial vendor. Regardless of sourcing, consumers are often incentivized monetarily to complete the questionnaire or test.

The advantages of on-line consumer surveys include the possibility of reaching a global population, very large sample sizes, flexibility of survey design, speed and timeliness of administration, ability to force response completion, ease of data acquisition/entry/analysis, ease of respondent completion, and low administrative costs (Evans & Mathur, 2005, 2018; Wright, 2005; Casler, Bickel, & Hackett, 2013; Callegaro, Manfreda, & Vehovar, 2015; de Winter, Kyriakidis, Dodou, & Happee, 2015). In addition, on-line questionnaires offer the advantage of greater perceived anonymity among respondents, encouraging more honest answers to sensitive questions than in-person surveys (Kreuter, Presser, & Tourangeau, 2008; Nikolaus, Ellison, & Nickols-Richardson, 2020).

However, as with any new data acquisition method, there are drawbacks. Prominent among these are the inability to provide clear one-on-one instruction to respondents, inherent sampling biases, self-selection bias, variability among respondents in their ability to access the survey due to device limitations, connectivity issues, privacy and security concerns, and generally lower response rates than other forms

* Corresponding author.
E-mail address: Sara.Jaeger@plantandfood.co.nz (S.R. Jaeger).

https://doi.org/10.1016/j.foodqual.2022.104676
Received 1 April 2022; Received in revised form 29 June 2022; Accepted 29 June 2022
Available online 3 July 2022

of survey administration (Wansink & Sudman, 2002; Wright, 2005). These factors, combined with the inability to control the questionnaire-taking environment (e.g., environmental distractions, background noise, multi-tasking, the presence of others or lack of attention due to boredom) can produce poor or unreliable data quality, or even worse, invalid data [see Evans and Mathur (2018) for a review and perspectives on the advantages, disadvantages, and evolution of online questionnaires].

1.2. Online data quality: An operationally defined construct

In the survey and questionnaire literature, data quality is generally recognized as a multi-dimensional construct (Lyberg, 2012), but one that lacks a standardized definition. Most often, it is defined operationally by the metrics used to assess specific dimensions of data quality. Thus, Vaughan (2019) defined data quality as a measure of the condition of data based on such factors as response accuracy, completeness, consistency, reliability and timeliness. Bowling (2005) defined it by the measured parameters of questionnaire response rates, item response rates, accuracy of responses, absence of bias, and completeness of information. Other researchers have included response consistency and respondent honesty (attention, effort) as additional critical dimensions of data quality (Moss & Litman, 2022). In its essence, data quality is a construct tied closely to the factors that influence the commonly accepted dimensions of accuracy, reliability, and completeness of responses to questionnaires, and it is operationally defined by the specific metrics used to assess these dimensions. As online questionnaires evolve in complexity and as modes of electronic administration advance, the operational definitions and metrics used to assess online data quality will undoubtedly evolve.

1.3. Online data quality: An often neglected element in questionnaire research

Despite the growing reliance on online questionnaires by sensory and consumer scientists and the potential for poor quality data due to uncontrolled elements of administration, published papers that utilize online questionnaire data infrequently report metrics of data quality or note the precautions that were taken to ensure valid data [for some recent exceptions, see Tonsor, Lusk, & Schroeder, 2022; Schifferstein, Lemke, & de Boer, 2022; Chong, Leung, & Qing, 2022]. This stands in stark contrast to the emphasis placed on obtaining high quality data from trained panelists in descriptive sensory analysis, where issues of panel reliability and discriminability are the focus of extensive research (e.g., McEwan, Hunter, van Gemert, & Lea, 2002; Bi & Kuesten, 2012; Tomic, Forde, Delahunty, & Næs, 2013; Sipos, Nyitrai, Hitka, Friedrich, & Kókai, 2021) and where panelist motivation is seen to be critical to data quality (Lund, Jones, & Spanitz, 2009). The fact that online consumer testing does not receive the same degree of scrutiny about these issues is curious, since it has been known for some time that respondents participating in online research often resort to satisficing behaviors, i.e. short-cuts and other convenience-based strategies, to minimize cognitive effort and time to complete the questionnaire conscientiously (Krosnick & Alwin, 1987; Krosnick, 1991; Krosnick & Berent, 1993; Krosnick & Fabrigar, 1997), which has been demonstrated to impact both consumer preference judgments, response consistency and willingness to pay estimates (Gao et al., 2016a,b; Jones, House, & Gao, 2015). In addition, with the evolution of crowd-sourced surveys and paid incentives, the source of respondent motivation is a growing issue (Eisenberger & Cameron, 1996; Heyman & Ariely, 2004; Stähli & Joye, 2016).

1.4. Aims and contributions

Against this background, the first two aims of the present paper are to 1) identify the most significant factors that can impact the reliability and validity of consumer data obtained using online questionnaires, and 2) highlight some of the available indices of online questionnaire data quality that can be used by researchers to assess the quality of their online data. The third aim, which directly builds on the first two, is to 3) recommend certain indices, countermeasures, and best practices that can be used by sensory and consumer researchers to assess and improve data quality obtained from online questionnaires.

The practical contribution of this paper is to alert researchers to the many factors that contribute to poor data quality and the specific data quality metrics and best practices that sensory and consumer scientists can use and report when conducting and publishing online questionnaire data. The top of Table 1 is a listing of these factors with directions to the specific section of this paper in which they are discussed. The bottom of Table 1 lists the various metrics of online data quality and where they appear in the text.

2. Factors affecting data quality in online consumer surveys

For the purposes of this paper, factors that can affect online data quality are categorized into three groups: 1) factors related to questionnaire design and administration, 2) factors related to respondent demographics and psychographics, and 3) factors related to distractions, carelessness, and maladaptive questionnaire-taking behavior.

2.1. Factors related to questionnaire design and administration

How a questionnaire is administered, in terms of the software platform that is used, the electronic devices on which it is completed, and rewards/ incentives offered to participants can all have important effects on data quality. In addition, as in all forms of questionnaires, the design elements of the questionnaire (e.g., vague or poorly worded questions, poorly constructed response choices or scales, excessive questionnaire length, etc.) can affect data quality.

Table 1
List of factors affecting data quality in online consumer questionnaires and the available metrics of online data quality. The right column is the section number in this paper in which the factor or metric is discussed.

Factors affecting data quality in online consumer surveys	Section
Questionnaire design and administration	
Mode of administration and survey device	2.1.1.
Questionnaire completion incentives	2.1.2.
Questionnaire topic/respondent interest	2.1.3.
Questionnaire length	2.1.4.
Item difficulty	2.1.5.
Item response modes and scales	2.1.6.
Respondent demographics and psychographics	
Respondent identity: human vs robot	2.2.1
Age and gender	2.2.2.
Respondent education and cognitive ability	2.2.3.
Respondent familiarity/literacy with data collection devices	2.2.4.
Respondent survey-taking experience	2.2.5.
Respondent personality characteristics	2.2.6.
Distractions, carelessness and maladaptive attitudes and behaviors	
Distractions in the questionnaire-taking environment	2.3.1.
Respondent attitudes toward questionnaires and surveys	2.3.2.
Careless responding	2.3.3.
Satisficing behavior	
Metrics of online data quality	
Ex-ante measures of data quality	
Instructional manipulation checks and trap questions	3.1.1
Bogus, infrequency, nonsense, and inconsistency questions	3.1.2.
Respondent engagement	3.1.3.
Post-hoc measures of data quality	
Questionnaire completion rates	3.2.1.
Questionnaire completion times	3.2.2.
Item response times	3.2.3.
Response accuracy	3.2.4.
Response quality	3.2.5.

2.1.1. Mode of administration and survey device

On-line questionnaires can be designed for execution on a variety of software platforms and electronic devices. The software platform, e.g. SurveyMonkey, Qualtrics, eQuestionnaire, Clicksurvey, ZipSurvey, or proprietary vendor platform, can impact data quality through limitations that the platform places on question formats, response modes, and display design/features. To the extent that the researcher is forced to shape their questionnaire to conform to the platform limitations at the expense of asking questions in the most theoretically sound manner, the overall quality of the data will be impacted.

As far as data quality is concerned, there is some evidence that crowd-sourced online questionnaires, e.g. Amazon's Mechanical Turk, may provide better quality than vendor panels (Zhang & Gearhart, 2020). The latter investigators found MTurk to produce better data on manipulation checks and higher completion rates. However, completion time was found to be shorter among crowd-sourced panelists, suggesting that crowd-sourced panels may be more likely to speed through the questionnaire in order to obtain incentives (Smith, Roster, Golden, & Albaum, 2016). Although the demographic information of MTurk respondents is not ascertainable in advance of survey execution, it has been found to be younger and more highly educated than the average population of vendor panels (Zhang & Gearhart, 2020), perhaps contributing to the observed differences.

In addition to format issues, different software platforms may cause the questionnaire to be displayed differently on different browsers and internet services. Depending on the quality of the survey display on the user's device/browser/internet provider, completion rates and data quality may vary (Fan & Yan, 2010). Thus, the researcher must adequately review and assess the appearance and functionality of their questionnaire on the software platform on which it is to be hosted. For this purpose, it is advisable to obtain screen shots or functional links of the questionnaires by device type and to keep these as study documentation.

Just as the software platform will affect questionnaire data quality, the electronic device(s) on which the questionnaire is or can be implemented will influence the quality of the obtained data. Until recently, respondents were restricted to and/or preferred to complete questionnaires on PCs (Toepoel & Ludtig, 2015; Revilla, Toninelli, Ochoa, & Loewe, 2016). However, software capabilities and preferences are changing, and the use of tablets and smartphones by respondents has a number of advantages for data acquisition. Among these are the ability to reach younger respondents who utilize smartphones regularly and the ability to collect questionnaire data in locations beyond the home or place of work/school. It has also been suggested that smartphone users are open to allowing other types of data to be collected, including installation of specific applications, taking photographs of test products or of the consumption environment, as well as providing GPS locations (Revilla et al., 2016). Newer smartphones also enable collection of biomarker and accelerometer data to provide measures of the physical and physiological conditions under which the data are obtained.

Although smartphones can increase participation rates, without appropriate changes to the questionnaire, formatting issues, such as inability to view the entire page, will arise. To avoid this, Mobile Browser Surveys (MBS), which allow access to the questionnaire on any electronic device, constitute the best approach for administration (Couper, 2000; Revilla, et al., 2014, 2016).

Disadvantages to using smartphones include the fact that questionnaires generally take longer to complete (Revilla et al., 2016), which may lower completion rates and increase dropouts. In addition, since the screen size is small, it limits the presentation of lengthy or matrixed response formats. The latter can be circumvented by the use of slider scales and graphic analogue scales that allow a position-and-point response for scaled response items (Toepoel, 2016). Buskirk and Andrus (2012) provide a summary of the strategies, advantages and disadvantages of smartphones for conducting questionnaires, while Tourangeau et al. (2018) provide a comprehensive analysis of potential data quality issues.

2.1.2. Questionnaire completion incentives

In general, completion incentives, which can take the form of cash payments, gifts, redeemable points, course credits, etc., dramatically increase completion rates in online questionnaires (Goritz, 2006; Becker, Möser, & Glauser, 2019), especially among less motivated respondents (Stähli & Joye, 2016). However, there is no direct relationship between the amount of the incentive and completion rates (Goritz, 2006; Porter & Whitcomb, 2003). Incentives also have been shown to improve the accuracy of willingness-to-pay (WTP) judgments in conjoint studies (Ding, Grewal, & Liechty, 2005). However, it has been noted that incentives may undermine intrinsic motivation (Eisenberger & Cameron, 1996; Heyman & Ariely, 2004). For example, several recent studies have shown that incentivizing respondents can produce increased levels of careless responding (Gadiraju, Kawase, Dietze, & Demartini, 2015; Niessen, Meijer, & Tendeiro, 2016; Shamon & Berning, 2020). This may result from incentives attracting less motivated respondents who are merely interested in obtaining the incentive but who employ satisficing behaviors to quickly complete the survey for this purpose. Berinsky, Margolis, and Sances (2016) found it difficult to increase the attentiveness of "shirker" respondents through incentives or other forms of motivation and, instead, argue for their identification through various forms of attention checks (see Section 3).

2.1.3. Questionnaire topic/respondent interest

The topic of the questionnaire and its interest to the respondent can drive questionnaire response rates and overall data quality (Groves, Presser, & Dipko, 2004; Maniaci & Rogge, 2014a; Liu & Wronski, 2018; Gummer & Roßmann, 2013; Roßmann, Gummer, & Silber, 2018). Maniaci and Rogge (2014) found that respondents who were internally motivated by the questionnaire (i.e., had a strong interest in the topic) were more attentive to the questionnaire than respondents who were only externally motivated (e.g., by incentives). However, highly sensitive questionnaire topics (e.g., food insecurity) can reduce respondent motivation and willingness to provide accurate information (Tourangeau & Smith, 1996; Kreuter et al., 2008; Roster, Albaum, & Smith, 2017).

2.1.4. Questionnaire length

A lengthy questionnaire can induce boredom, foster a lack of attention, and negatively affect completion rates (Cannell & Kahn, 1968). Respondents often start a questionnaire but drop out when it becomes too long (Saxon, Garratt, Gilroy, & Cairns, 2003; Ganassali, 2008; Galesic & Bosnjak, 2009; Liu & Wronski, 2018; Andreadis & Kartsouniduo, 2020). In general, questionnaires that take 15–30 min to complete result in higher drop-out rates than those that only take 5–10 min (Deutskens, De Ruyter, Wetzels, & Oosterveld, 2004; Galesic & Bosnjak, 2009; Mavletova, 2013). In one study that asked respondents their ideal length of time for completion of an online questionnaire, their recommended median length was 10 min, with a maximum of 20 min (Revilla & Ochoa, 2017). Longer questionnaires have also been shown to increase missing data (Peytchev & Tourangeau, 2005; Galesic, 2005; Galesic & Bosnjak, 2009), satisficing behaviors (see Section 2.3.4) (Chen, 2011; Barge & Gehlbach, 2012; Guidry, 2012) and the proportion of "don't know" responses (Deutskens et al., 2004).

The length of product preference questionnaires can be of particular concern. Morii, Sakagami, Masuda, Okubo, and Tamari (2017) conducted a preference survey of 100 visual stimuli in which respondents indicated their degree of preference for each stimulus using a line scale. Across several studies, the authors found that mean preference ratings increased over the course of the first 20 stimuli, reaching a stable level before declining. Furthermore, maximal and minimal preference ratings were observed in the early trials of the survey, suggesting that early parts of the questionnaire generated better discriminability among stimuli, but as the questionnaire progressed, differential responding

decreased, perhaps due to increased boredom.

To circumvent the problems posed by lengthy questionnaires, some investigators remove items that are part of embedded standardized scales, despite the potential of this to compromise the validity of the questionnaire (e.g., Smith, McCarthy, & Anderson, 2000; Maloney, Grawitch, & Barber, 2011). From the authors' perspectives as journal editors, this approach to shortening questionnaires is often seen in consumer studies of food behavior, where validated state, trait or other standardized scales [e.g., the Food Neophobia Scale (Pliner & Hobden, 1992) or the Food Choice Questionnaire (Steptoe, Pollard, & Wardle, 1995)] are shortened by choosing the "most relevant" scale items and deleting those considered less relevant. A better strategy is to use a split questionnaire design in which a long questionnaire is divided into two shorter questionnaires with each administered to a different sample of respondents (e.g., Johnson, 2005; Adigüzel & Wedel, 2008; Fricker et al., 2012; Huang, Curran, Keeney, Poposki, & DeShon, 2012). In one study in which a long questionnaire was divided into two smaller ones, both response rates and data quality (percent of mid-point responses and completeness of open-ended questions) were found to be better in the split questionnaire (Andreadis & Kartsounidou, 2020). However, two drawbacks of a split-questionnaire are that it reduces the number of responses to any one item and eliminates within-subject comparisons across items appearing in the different sections.

2.1.5. Item difficulty

Yan and Tourangeau (2008) found that both the number of clauses in a questionnaire item, the number of words per clause, and the number of response alternatives had a detrimental effect on response time due to the cognitive burden placed on the respondent, resulting in longer processing times. These investigators also found that respondents were generally faster with factual questions than with attitudinal questions, which is an important distinction when conducting food-related surveys. In fact, Subar et al. (2001) found that the clarity of the instructions and items in a traditional food frequency questionnaire can compensate for its overall length. Also, reverse-worded questions (those in which the meaning and direction of the sentence is reversed through the use of a negative or antonym) can slow response times (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003) and interfere with the factor structure of the scale (Marsh, 1996). Reverse-worded items can also introduce response variability, especially in cross-cultural surveys, where interpretation of the meaning of negatively worded statements may vary (Wong, Rindfleisch, & Burroughs, 2003). In a review of the use of reverse-worded questions, Weijters and Baumgartner (2012) supported their use, but recommended they appear throughout the questionnaire (not just as a "catch" item) and that they be "polar opposite reversals," i.e. restated to reflect a polar opposite opinion and not a simple negation, i.e. using a negative word to deny the original proposition. The reader is referred to Weijters and Baumgartner (2012) for a detailed discussion.

2.1.6. Item response modes and scales

In sensory and consumer psychology, there is a large literature discussing the most appropriate response scales to use when obtaining quantitative judgments (see Lim, 2011: Jaeger & Cardello, 2009; Louviere, Flynn, & Marley, 2015 for reviews). In other areas of psychology and marketing, the emphasis has focused primarily on best scale lengths for simple Likert-type or category scales (Cox, 1980; Matell & Jacoby, 1971; Niemi, Carmines, & McIver, 1986; Krosnick & Fabrigar, 1997; Lee & Paek, 2014) and on consumer "response styles" that predispose certain types of ratings (Baumgartner & Steenkamp, 2001). Within the latter literature, an often addressed issue is whether a verbal or numeric scale is the most appropriate for use in questionnaires. For example, Saris and Gallhofer (2007) found that category response alternatives that were verbally labeled had higher reliability than simple numeric scales. Similarly, Weng (2004), Menold, Kaczmirek, Lenzner, and Neusar (2014) and Menold (2020) found that scale reliability was greater when all response alternatives on a category scale were verbally labeled, as opposed to scales in which only the end-categories were labeled. The latter investigators used eye-tracking to show that the verbal scales were superior in reliability and validity to numeric scales, due in part to longer fixation times and the required time to endorse numeric responses. In terms of validity, Windschitl and Wells (1996) also found verbally labeled category response alternatives to be better than numeric ones. Other studies showing verbal category response modes to be better than numeric responses include Toepoel and Dillman (2011) and Tourangeau, Couper, and Conrad (2007), while Krosnick and coworkers (Krosnick & Alwin, 1987; Krosnick & Berent, 1993; Krosnick & Fabrigar, 1997) found that verbal response scales produced less satisficing behavior.

With regard to other response modes, Toepoel and Funke (2018) found better responding on mobile questionnaires with the use of radio buttons, as opposed to slider scales. Visual analogue scales with a point and click option also performed better than did the sliders. As an alternative to verbal, numeric and analogue response formats, the use of emoji in online questionnaires has been shown to increase response speed and questionnaire enjoyment, especially among millennials (Bosch & Revilla, 2021).

2.2. Factors related to respondent demographics and psychographics

Among the demographic and psychographic aspects of respondents that have been assessed for their effect on online data quality are age, education, cognitive ability, computer literacy, survey-taking experience and personality variables. Although the demographics of panel populations are relatively easy to control through selection quotas, the psychographic or personality factors that predispose a respondent toward better or poorer responding are usually unknown and cannot be assessed unless measured as part of the questionnaire itself. In addition, in today's computer testing environments, there is a risk that data are being generated by artificial intelligence (AI) and not by a human respondent.

2.2.1. Respondent identity: Human vs robot

Respondent identity refers to the ability to verify who is completing an online questionnaire. Questionnaires conducted using commercial vendors and pre-established panels of respondents can verify the identity of the respondent and link their survey data to demographic or other data on file for that respondent. However, for many online questionnaires, the identity of the respondent is unknown. In addition, there is the growing problem of AI bots masquerading as human respondents. Such bots can perform automated tasks like completing online questionnaires to obtain paid incentives (Dupuis et al., 2020). In one recent online posting about an eating behavior questionnaire (Simone, 2019), the author stated that only 3% of questionnaires could be eliminated as *not* coming from an AI bot, with one indicator being the number of surveys in which all questions were responded to in an identical manner. To address this growing problem, a variety of techniques have been developed to counteract such bot attacks (Storozuk, Ashley, Delage, & Maloney, 2020). Among these are CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) questions, which can be administered as part of the questionnaire. These questions present a set of distorted letters and numbers or images on the device screen, which the respondent must read, decipher, and accurately respond to, in order to proceed with the questionnaire. In a recently published on-line questionnaire focusing on food package designs, Schifferstein et al. (2022) included such a CAPTCHA question to eliminate bots from the study. The percentage of failures on the CAPTCHA question was not reported.

2.2.2. Age and gender

Although some investigators have found that older individuals are less likely to participate in online questionnaires (Moore & Tarnai, 2002), others have failed to find such an effect of age (Nielsen, 2011;

Lindhjem & Navrud, 2011). However, Yan and Tourangeau (2008) found increased item response times among older respondents and Harms, Jackel, and Montag (2017) found that older respondents required more time to complete a questionnaire. These effects may be related to confounding variables, such as device familiarity or simple motor reactivity. For example, Fricker, Galesic, Tourangeau, and Yan (2005) showed that, while older respondents were slower to complete online surveys, the difference was not as large in telephone surveys. This is not to say that age itself is not a factor. Andrews and Herzog (1986) found older respondents to have higher variability in questionnaire responses, even for traditional paper and pencil formats. Further, it has been found that the removal of respondents who fail attention checks in surveys can produce sample bias due to removal of many older respondents (Vannette, 2016). In contrast to the above findings, Maniaci & Rogge (2014) found less inattention among older respondents, suggesting that older respondents may have fewer distractions when completing questionnaires than younger respondents, especially if the questionnaire fails to capture the interest of the younger respondent. With regard to data quality and attention, a multi-national study (Gao, et al., 2016a) found older respondents (and higher income respondents) to be less likely to fail instructional manipulation checks in which a directed response is required.

Regarding gender, although there is evidence that females are less frequent and less intense users of the internet (Ono and Zavodny, 2003, 2016), they are more likely to participate in surveys of all types than males (Curtin, Presser, & Singer, 2000; Singer, Van Hoewyk, & Maher, 2000; Moore & Tarnai, 2002; Smith, 2008). For this reason, setting quotas for gender is necessary to avoid a preponderance of females, especially in food-related surveys, due to topical interest and the traditional role of females as primary food shoppers. However, neither Olsen (2009), Lindhjem and Navrud (2011) nor Nielsen (2011) found differences in gender representation in online vs other forms of questionnaires. So too, the evidence is weak on any differences between males and females in data quality. Although Maniaci and Rogge (2014b) found better attention to online questionnaires among female respondents, neither Malhotra (2008) nor Polonsky and Vocino (2010) found effects of gender on data quality indices.

2.2.3. Respondent education and cognitive ability

A variety of studies have shown that more educated or affluent individuals are more likely to participate in surveys than less educated/affluent individuals (Curtin, Presser, & Singer, 2005; Goyder, Warriner, & Miller, 2002; Nielsen, 2011). Regarding questionnaire data quality, Maniaci and Rogge (2014) found greater attention among more highly educated respondents, while Yan and Tourangeau (2008) found high school graduates to have faster response times on internet questionnaires than individuals who did not complete high school. Similarly, Lindhjem and Navrud (2011) found a greater use of the "don't know" category among less educated respondents. These studies stand in contrast to others by Galesic (2006) and Harms et al. (2017), who found no relationship between questionnaire dropout rates or response speed and education level. In general, the data on education level and data quality, like those for age and gender, are mixed, perhaps due to the interaction of education level with other sociodemographic and interest-based variables (e.g., age, affluence, computer literacy, topic interest, interest in questionnaire taking).

Respondent cognitive abilities are rarely considered in research on online data quality, except in cases where the targeted population is known or suspected to have cognitive impairments. Most often, this arises in research with the elderly or those in nursing homes and other public or private institutions of the infirm. However, in one recent study, Seelye et al. (2018) found completion times for online questionnaires to be positively correlated with increasing cognitive impairment. In still another study of nursing home respondents, a higher level of item non-responsiveness was found with increasing cognitive impairment as measured by clinical indices. Those individuals with no or mild impairment were relatively adept at completing questionnaires, but those with moderate or high cognitive impairment produced poor data quality (Kutschar, Weichbold, & Osterbrink, 2019).

2.2.4. Respondent familiarity/literacy with data collection devices

It is well known that there are generational differences in computer and electronic device literacy, as well as between people with greater or lesser education. Even simple devices can show an influence of experience, as shown by Toepoel (2016), who found that respondents with greater experience in completing phone-based questionnaires evaluated their experience better than did respondents with less experience. Similarly, Yan and Tourangeau (2008) found that individuals with greater experience using the internet had faster response times in online questionnaires than did respondents with less experience. In general, respondents who have greater experience with the internet, computers or electronic devices produce better online data quality.

2.2.5. Respondent survey-taking experience

Although using vendor or institutional panels can ensure quotas and respondent identities, it has been estimated that as many as 60 percent of vendor panelists belong to multiple vendor panels (Willems, Van Ossenbruggen, & Vonk, 2006). In fact, one study found that vendor panelists belong to an average of four different panels (Gittelman & Trimarchi, 2010). Panelists who participate in numerous online questionnaires over a period of time are colloquially referred to as "professional respondents" (Hillygus, Jackson, & Young, 2014). The designation of a consumer as being a professional respondent varies in the literature but ranges from 4 to 30 questionnaires per month (Honda & Motokawa, 2005; Gittelman & Trimarchi, 2010; Frede, Markowitz, & Coffery, 2006). It has been documented that the primary motivations for such respondents are the incentives, either monetary or in kind (Gittelman & Trimarchi, 2010; Mason & Watts, 2009).

The most significant issue of data quality with respondents who have substantial experience in completing questionnaires is that they often speed through the questionnaire (Knapton & Garlick, 2007; Toepoel, Das, & Van Soest, 2008; Yan & Tourangeau, 2008; Hillygus et al., 2014) or answer questions strategically so that they do not have to answer follow-up questions (Nancarrow & Cartwright, 2007), both of which are forms of satisficing behavior. On the positive side, frequent responders tend to give fewer neutral or "don't know" responses (Smith & Brown, 2006; Binswanger, Schunk, & Toepoel, 2013) and have higher consistency on related questionnaires (Garland, Santus, & Uppal, 2009). Other positive elements of these respondents are that they have a more positive attitude toward questionnaire-taking (De Wulf & Berteloot, 2007) and are more honest in responding to demographic questions, e.g., race and income (Smith & Brown, 2006). Taken together, the findings on data quality among frequent questionnaire takers are mixed, with both positive and negative elements [see Sandorf, Persson, and Broberg (2020) for a recent analysis of the effect of professional respondents on data quality.].

Perhaps most important to sensory and consumer scientists are the differences between frequent and less frequent questionnaire takers in their overall response to products. The more questionnaires that a respondent completes, the less nominal they become (Chang & Krosnick, 2009). For example, there is evidence that frequent questionnaire takers have higher purchase interest for products (Walker, Pettit, & Rubinson, 2009) and are also less impressed by new products than unexperienced consumers (Miller, 2007). Despite these findings, other researchers have failed to find significant differences between frequent questionnaire takers and nominal populations in terms of attitudes and behaviors (Pineau, Nukulkij, & Tang, 2005; Kruse et al., 2009).

2.2.6. Respondent personality characteristics

As far as the personality characteristics of respondents are concerned, Maniaci and Rogge (2014b) administered the 44-item Big Five Inventory of personality to respondents across multiple questionnaires.

Results showed that respondents who were attentive to the questionnaire were more likely to be conscientious, agreeable, and open to new experiences. In addition, respondents who had higher self-esteem and socially desirable response tendencies showed less careless responding. Similarly, Bowling et al. (2016) found conscientiousness, agreeableness, extraversion and emotional stability to produce less inattention on questionnaires. Most recently, Robinson and Boies (2021) found that respondents who were more conscientious, agreeable, and open also produced better data quality than those who were not.

Another personality trait familiar to sensory and consumer scientists is Need for Cognition (NFC), which reflects the extent to which individuals are inclined towards effortful cognitive activities (Cacioppo & Petty, 1982; Cacioppo, Petty, & Kao, 1984). High NFC respondents have been shown to be less likely to fail instructional manipulation checks (see Section 3.1.1; Oppenheimer, Meyvis, & Davidenko, 2009). Similarly, Hauser and Schwarz (2015) showed that Cognitive Refection, a measure of an individual's predisposition to engage in reflective and systematic thought (Frederick, 2005), was associated with better performance on attention measures. They showed that Cognitive Reflection scores and trap question performance were most highly associated when the trap question occurred early in the questionnaire. When the trap question occurred later in the questionnaire, no relationship was observed. They interpreted the relationship to be due to the trap question cuing the respondent that more reflective thinking would be required to successfully answer questions on the survey, which was associated with higher scores on Cognitive Reflection.

As opposed to the above personality *traits*, mindfulness is a psychological *state* defined as "open or receptive awareness of and attention to what is taking place in the present" (Brown & Ryan, 2003). As a psychological state, it can be achieved through meditation and mindfulness training in which the individual is taught to ignore distractions and to focus on their awareness of details of the environment and the task before them. Although the relationship of mindfulness to online data quality has not been fully examined, MacLean et al. (2010) showed that mindfulness meditation training can increase perceptual sensitivity and attention, while Mrazek, Franklin, Phillips, Baird, and Schooler (2013) showed that a short mindfulness training task significantly reduced mind wandering in a sustained attention task. Similarly, in an unpublished study, Cardello and Schutz (2010) introduced a short mindfulness induction treatment with consumers participating in a CLT test of food likes/dislikes. The induction group had better discrimination among products than did the non-induction group. As such, mindfulness training, which can be accomplished online at the start of a questionnaire, may be considered as a proactive tool to improve online data quality, and the measurement of respondent mindfulness using available mindfulness measures (Brown & Ryan, 2003; Bohlmeijer, Ten Klooster, Fledderus, Veehof, & Baer, 2011) should be examined for its relationship to measures of online data quality, such as those outlined in Section 3.

2.3. Factors related to distractions, carelessness and maladaptive attitudes and behaviors

2.3.1. Distractions in the questionnaire-taking environment

Regardless of instructions to the contrary, panel respondents may complete surveys under highly distracting conditions. These may involve completing the survey in a noisy environment, while listening to music or streaming movies/series, while social networking, gaming, or on-line shopping, while in the presence of or interacting with other individuals, completing the questionnaire concurrently with other work tasks, while emailing, etc. In addition, if questionnaires are allowed to be completed on mobile devices, the number and magnitude of environmental distractions can be more severe, including completion of the questionnaire in bars or restaurants, at sporting events, or in any number of highly distracting locations outside of the home. As described by Tourangeau, Rips, and Rasinski (2000) environmental distractions can result in respondents who do not adequately engage in one or more of a number of fundamental stages of cognitive processing: (1) accurately interpreting the meaning of the survey item; (2) searching memory for the relevant information; (3) summarizing this information into a unified judgment; and (4) conveying that judgment in the most accurate way, given the available response options.

2.3.2. Respondent attitudes toward questionnaires and surveys

Poor attitudes toward questionnaires and survey taking can negatively influence data quality (Sjoberg, 1954), including response rates (Baruch, 1999). Rogelberg and co-workers (Rogelberg, Luong, Sederburg, & Cristol, 2000, 2001) examined respondent attitudes toward questionnaires and questionnaire-taking and developed a measure of those attitudes. Their measurement construct involves two aspects of these attitudes –enjoyment and value, each measured by three questionnaire items. Items measuring survey/questionnaire enjoyment included such statements as "surveys are fun to fill out" and those designed to capture inherent value included such statements as "a lot can be gained from information gathered from surveys." Using this measure, Rogelberg, Fisher, Maynard, Hakel, and Horvath (2001) found that high levels of enjoyment and perceived value produced fewer missing responses on open-ended questions, higher rates of following directions accurately, and a greater willingness to participate in future questionnaires. Higher enjoyment alone produced faster timeliness of completion.

2.3.3. Careless responding

In addition to distractions and poor attitudes, carelessness can impact the quality of data obtained from online questionnaires. This can result from lack of interest in the survey topic (Anduiza & Galais, 2017; Gummer, Roßmann, & Silber, 2021), poor survey incentives, or the anonymity of surveys that allows less motivated respondents to be less conscientious when completing the survey (Douglas & McGarty, 2001). Careless responses are a ubiquitous problem, with estimates of careless responding varying from 33% (Goldammer, Annen, Stöckli, & Jonas, 2020) to 46% (Oppenheimer et al., 2009), and as high as 78% in one recent study (Mancosu, Ladini, & Vezzoni, 2019). Questionnaire respondents who are careless are often referred to as "careless responders" (Meade & Craig, 2012), "insufficient effort responders" (Huang et al., 2012) or "careless/ insufficient effort (C/IE) responders" (Curran, 2016). Regardless of cause, careless responding is an important source of invalidity in online questionnaire research (Huang, Liu, & Bowling, 2015), and eliminating careless respondents has been shown to be necessary to increase effect sizes (Brühlmann, Petralito, Aeschbach, & Opwis, 2020).

2.3.4. Satisficing behavior

Satisficing is a type of behavior often exhibited by respondents in both online questionnaires and other forms of sensory and consumer testing. Simon (1957) and Simon, Stedry, Lindzey, and Aronson (1968) first used the concept to explain consumer decision-making behavior that did not result in maximal gain or optimal responding, but more recently it has been applied to explain a variety of undesirable or suboptimal responses among individuals participating in consumer tests (Tourangeau, 1984; Krosnick, 1991). In this context, the construct of satisficing refers to the fact that some respondents only invest enough cognitive or physical energy to make a "satisfactory" response decision, rather than one that is the best reflection of their true knowledge, attitude, or opinion, given their factual knowledge and inherent ability to respond accurately and conscientiously. Often, such consumers do not invest sufficient time or attention to reading and fully comprehending what is being asked or to carefully assess which of several response options most accurately reflects their attitude or opinion. They often respond with "safe" answers or those that are the easiest to make.

Krosnick (1991) identified several characteristics of questionnaire responses that can indicate satisficing behavior. Among these are selecting the first response option that seems reasonable, agreeing with

assertions in the survey or endorsing the status quo, rushing through survey items, randomly selecting responses or not differentiating among response options, i.e. straight-lining (Greszki, Meyer, & Schoen, 2014; Zhang & Conrad, 2014; Schonlau & Toepoel, 2015). Other forms of satisficing behaviors include selecting "don't know," "not applicable" or mid-point (neutral) responses (Blasius & Thiessen, 2001), skipping items entirely, or dropping out of the survey prior to completion. Preference surveys that use like/dislike responses with a middle or neutral (neither like nor dislike) category are particularly prone to satisficing behavior, because they allow respondents to make safe or noncommittal responses, using these mid-point categories. Another common form of satisficing behaviors is often found when using rating scales where the consumer rounds to the nearest "10," only uses numbers that are easily divisible by the numbers "5" or "10" (Holbrook et al., 2014), or places hashmarks only near labeled points on graphic rating scales (Lawless, Sinopoli, & Chapman, 2010). In one analysis of the extent of satisficing behavior among European and U.S. survey respondents, Barge and Gehlback (2012) found that as many as 25% of respondents exhibited one of four different types of satisficing behaviors examined in their study. Moreover, these investigators found significant differences in responding to questionnaire items between individuals categorized as exhibiting or not exhibiting satisficing behavior.

3. Measures of online data quality: *ex-ante* and *post-hoc* indices

Regardless of whether data quality is compromised by factors related to the survey design or administration, respondent demographics or psychographics, or distractions, attitudes and maladaptive questionnaire-taking behaviors, a number of *ex-ante* (before the event) and *post-hoc* (after the event) indices can and should be used by researchers to provide a gauge of the quality of their questionnaire data (Johnson, 2005; Curran, 2016; Curran & Hauser, 2019; Shamon & Berning, 2020).

Further, regardless of the data quality metrics employed, authors should include a prominent and detailed statement in their manuscripts that describes the specific type of data quality protocols and metrics that were used, the results of the metrics (in terms of the percentage of respondents who demonstrated good/poor data quality and the criteria used), as well as the specific actions that were taken regarding respondents who did not pass data quality criteria [see Neissen, et al. (2016) regarding the necessity of the above actions]. Such protocols and metrics should be employed and reported, regardless of whether the online study is conducted as part of methodological research or routine, practical studies in sensory/consumer evaluation. It should also be kept in mind that each study will have a different set of factors that could reduce data quality and that will require a different set of protocols and data quality indices to best address and/or quantify the impact of those factors. If data quality protocols and metrics are not used, authors should provide justification for not doing so.

3.1. *Ex-ante measures of data quality*

A wide variety of indices of data quality have been developed that seek to identify inattentive or careless respondents through the introduction of specific items into the questionnaire before the questionnaire is administered. Such *ex-ante* measures take a variety of forms, as outlined in the sections that follow.

3.1.1. *Instructional manipulation checks and trap questions*

Attention check items, also known as "trap questions," consist of questions inserted into the survey followed by text that instructs the individual to respond with a specified response. Oppenheimer et al. (2009) developed a specific type of attention check item known as the Instructional Manipulation Check (IMC). The IMC is a statement or item in the survey that informs the respondent to ignore the standard response format and to, instead, provide a confirmation that they have read the manipulation by responding in a specified manner. An example of IMC is the blue-dot task shown in Fig. 1a (Oppenheimer et al., 2009) . A generic form of the IMC that can be used within a long series of questions with identical response formats is shown as the final statement in Fig. 1b.

Oppenheimer et al. (2009) included their IMC as part of a survey administered to respondents who were motivated by the topic of the survey and a group who were not motivated by the topic. The failure rate to the instructional manipulation check was significantly higher in the unmotivated group than the motivated group. In other studies, Oppenheimer et al. (2009) showed that eliminating respondents who failed the instructed manipulation check enabled the identification of significant differences in data sets that were found not to be significantly different without the elimination of these individuals. In particular, it was demonstrated that those respondents who had a high failure rate on the IMC also spent significantly less time on the survey than did respondents who correctly answered the IMC. Similarly, Liu and Wronski (2018) demonstrated that individuals who failed such trap questions were more likely to straight-line responses and to have lower consistency of responding than those who passed these questions. Finally, Gao, et al. (2016a), in a study of willingness to pay for apples, found that on-line questionnaire respondents who passed an instructed manipulation check had more efficient and more accurate WTP (willingness to pay) estimates than individuals who failed the manipulation check.

Although trap or instructed manipulation questions are often placed randomly within the survey, placing these items early in the survey makes it more likely that they will be read in full by the respondent and can improve the overall quality of the data. As such, they can serve as proactive techniques to improve respondents' attention (Hauser & Schwarz, 2015) and not simply measures of inattention. However, as a note of precaution, these authors also warned that doing so affects responding on the substantive portion of the survey and so, should be used with caution.

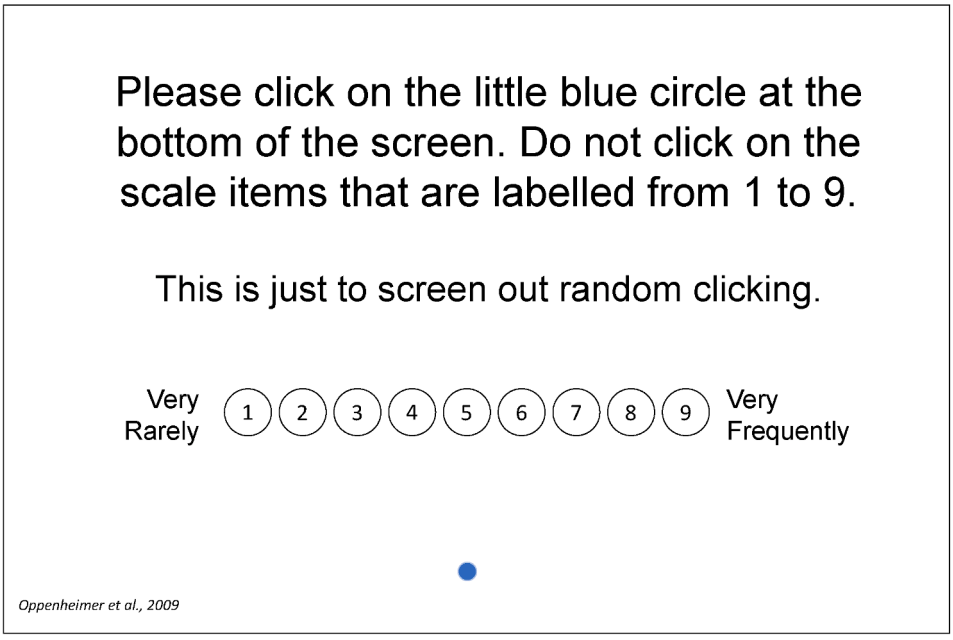
3.1.2. *Bogus, infrequency, nonsense, and inconsistency questions*

Bogus or infrequency questions are those for which a positive response is impossible or highly unlikely during honest or attentive responding, e.g. "I have traveled to every country in the world" or "I was born on February 30th" (Beach, 1989; Meade & Craig, 2012; Maniaci & Rogge, 2014b; Huang et al., 2012, 2015). Positive responses to these questions raise concerns about the validity of the respondent's other answers and/or the degree to which that respondent is paying attention to the survey items. Multiple bogus questions can be placed within a single survey and the sum of positive responses can be used as an index of inattention or invalidity.

In a recent study, Curran and Hauser (2019) examined the likelihood of false positives to a large number of infrequency measures. They also examined whether in cases of false positives, the survey respondent had a valid or viable rationale for their response. Although the sample size in their research was relatively small, their results showed that the best performing infrequency measures were those found in the often-used scales of Meade and Craig (2012) and Huang, Bowling, Liu, and Li (2014), and the authors recommended the use of published and successful infrequency measures, as opposed to creating new or novel ones. In addition, the authors found that in several cases, false positives to infrequency measures were, subsequently, well justified by the respondent, often-times showing a high degree of logic and rationalization. Although strongly supporting the use of infrequency measures as a check for careless/insufficient effort (C/IE) responding, the authors cautioned that the researcher must be diligent and conscious of the fact that, in some cases, they may be eliminating data of conscientious responders (Curran & Hauser, 2019).

A classic example of a nonsense question was used by Meiselman, Waterman, and Symington (1974) in a study of military food preferences that contained a long list of food names for which subjects were asked to indicate their liking or disliking for the food or to give a "don't know"

A)



B)

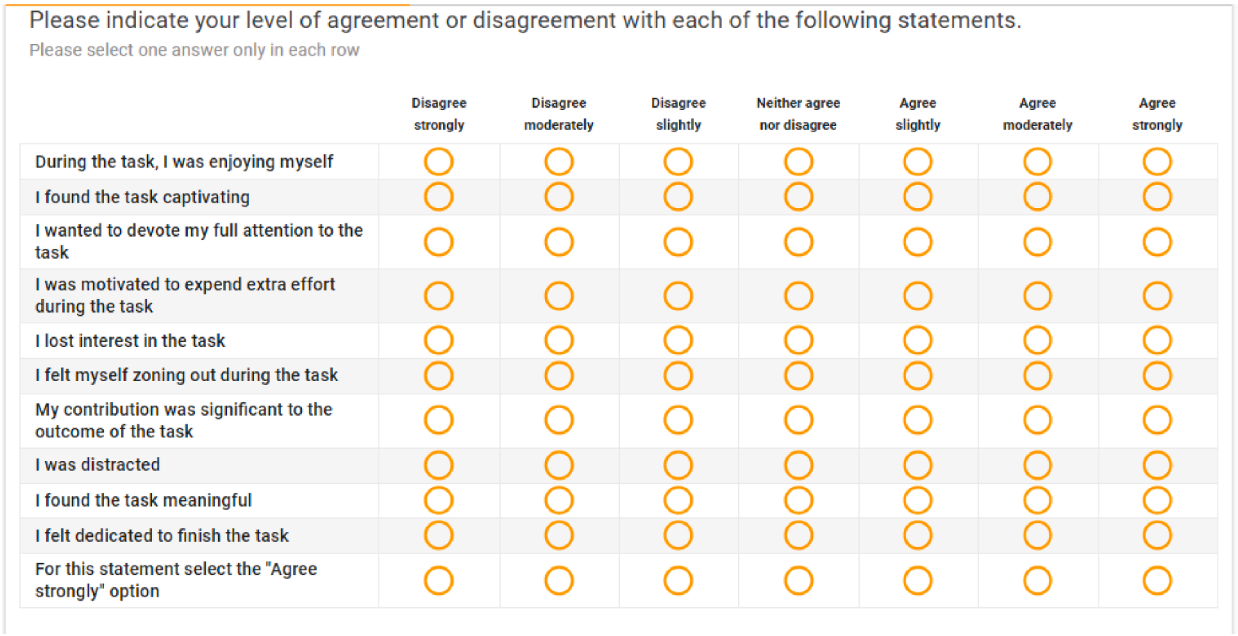


Fig. 1. Two Instructional Manipulation Check (IMC) examples.

response if they were not familiar with the food items. The researchers included several nonsense items in the survey (e.g., *funistrada*, *braised trake*, and *buttered ermal)*. If respondents rated these fake food names as anything other than "don't know," it was a basis to suspect that the respondent was not carefully reading and/or conscientiously responding to the items in the survey. Other types of nonsense questions involve the use of Likert statements of agreement/disagreement. Statements like "My local food store uses leprechauns as delivery boys" or "I often eat hamburgers made from ground, lean unicorn meat" are used at various points in the survey. Survey data from respondents who agree with such statements can be flagged for deletion from the data set with the assumption that such individuals are not attending to survey questions.

Inconsistency or reverse questions are another *ex-ante* technique that can be used as an attentional measure in online surveys. These consist of two paired questions in a survey that express completely opposite attitudes or require completely opposite answers, e.g. "I like carrots" and "I dislike carrots." The degree of inverse correlation between responses on such questions can be taken as an index of inattention or invalidity of survey responses.

3.1.3. Respondent engagement

In the on-line survey literature, the term "engagement" is used to refer to the respondent's ability to focus attention and become fully involved in the online questionnaire and its purpose (Laurel, 1993; Hull & Reid, 2003). Research on the elements of online questionnaires that make them engaging has found the attributes of challenge, positive affect, endurability, aesthetic and sensory appeal, attention, feedback, variety/novelty, interactivity, and perceived user control as important elements of such questionnaires (O'Brien & Toms, 2005, 2008). In order to quantify such engagement, the latter investigators developed a 31-item "User Engagement Scale" (O'Brien & Toms, 2010) tailored to on-line activities.

In the sensory and consumer area, engagement is considered essential to obtaining high quality data in both CLT tests and in survey research. In CLT tests, the sterile environment of sensory test booths and the presentation of numerous samples with large numbers of accompanying questions can lead to boredom and result in a loss of engagement. Bangcuyo et al. (2015) highlighted this problem as part of research focused on providing a more immersive environment for sensory testing in which consumers could better focus on the tasks at hand through a virtual reality environment. In their research, Bangcuyo et al. (2015) developed a 21-item engagement questionnaire based on the User Engagement Scale of O'Brien and Toms (2010) that addresses 8 different factors important to user engagement. This questionnaire was found to be sensitive to differences between a traditional testing environment and one considered more immersive and contextually relevant, with higher engagement ratings found in the more immersive environment. Further, the differences maintained themselves in repeat testing over the course of a month. Hathaway and Simons (2017) used a 19-item version of the Bangcuyo et al. (2015) questionnaire in another test of immersive environments and, again, found significant differences between the testing conditions employed in their research.

The initial work on engagement questionnaires in sensory and consumer research was followed by research by Hannum (Hannum & Simons, 2020; Hannum, Forzley, Popper, & Simons, 2020), who sought to develop a new scale of engagement, because the earlier scales were constructed in such a way as to jeopardize the validity of the dimensionality of the original scales upon which they were based. Hannum and Simons (2020) developed a three-factor (active involvement, purposeful intent, affective value) 10-item Engagement Questionnaire (EQ) specific to the purpose and environments common to sensory and consumer science. Each of the scale items in the EQ is rated on a 7-pt Likert scale, with three items reversed for scoring purposes. The first 10 items in Fig. 1b are those used in the EQ.

Hannum and Simons (2020) validated their scale in a variety of sensory and consumer CLT tests, and, as with the studies by Bangcuyo et al. (2015) and Hathaway and Simons (2017), the authors also validated the EQ in a study of immersive environments, showing that more immersive environments produced greater engagement, as reflected in higher EQ scores when the questionnaire was presented after completion of the sensory test (Hannum et al., 2020). Although not employed to date to assess subject engagement in online questionnaires, the EQ has the potential to serve as a valid and reliable test of user engagement in such surveys, because the construct of engagement, like mindfulness, is centered on focus and attention to task performance. As such, it would be predicted that individuals who score high on the EQ would produce higher quality survey data, i.e., fewer instances of satisficing, careless or inattentive responses.

As it relates to improving engagement in online surveys, focus has been placed on visual enhancement of surveys, with color and graphics, as well as using alternative response modes that include slide bars or drag and drop response modes, rather than radio buttons (Downes-LeGuin, Baker, Mechling, & Ruyle, 2012). Recently, gamification (incorporating game-like features into surveys) has also been proposed as a method to increase respondent engagement (Harms, Wimmer, Kappel, & Grechenig, 2014, 2015; Keusch & Zhang, 2017). Downes-LeGuin et al. (2012) proposed a survey taxonomy with four levels of increasing engagement by respondents. In increasing order these survey types are 1) text only, 2) decoratively visual, 3) functionally visual, and 4) gamified. In recent research comparing traditional online surveys to gamified surveys, Triantoro, Gopal, Benbunan-Fich, and Lang (2019) showed gamified questionnaires to produce greater affect, as reflected through an enjoyment survey administered to respondents as well as improved attention (Triantoro, Gopal, Benbunan-Fich, & Lang, 2020). The reader is referred to these latter works for a review of the opportunities offered by gamified questionnaires.

3.2. Post-hoc measures of data quality

As opposed to *ex ante* measures, many of which have been developed only with the advent of online questionnaires, *post-hoc* measures have been used for many years, going back to when mail and telephone surveys were the predominant forms of questionnaire administration. Nevertheless, these indices are still important to assess overall data quality of online questionnaires.

3.2.1. Questionnaire completion rates

Questionnaire completion rate refers to the proportion of respondents who complete the entire questionnaire divided by the number of respondents who either received the questionnaire or started the questionnaire but did not complete it. Completion rates are important indices of data quality in most on-line questionnaires, because they determine the total number of questionnaires for which the researcher has data and can affect the cost of conducting the survey (low completion rates require a greater number of respondents to be recruited in order to meet population quotas). Completion rates can also introduce sample bias into the research, because respondents who complete the questionnaire may be different in some way (based on demographics or psychographics) from those who fail to complete it, meaning that the survey sample may reflect only people who are highly engaged in the topic or motivated by the incentives [see Hays, Liu, and Kapteyn (2015) for a discussion of the factors affecting questionnaire completion rates].

3.2.2. Questionnaire completion times

Questionnaire completion time is another common measure of data quality because it speaks to the degree of attention or inattention that the respondent pays to the questionnaire items. Exceedingly long completion times may indicate inattentive respondents. Exceedingly short completion times may reflect the fact that the respondent did not read the questions carefully, perhaps responding randomly or in a straight-line manner to finish the questionnaire quickly. Extremely short completion times may also indicate automated responding, e.g., by an AI bot. Although the minimum completion time for a questionnaire below which a respondent can be considered a "speeder" has been defined as 60% of the median time to completion across respondents (Roßmann, 2010), a variety of more complicated indices have also been developed (Greszki, Meyer, & Schoen, 2015). The reader is directed to Goldammer et al. (2020) for a recent analysis of completion times and careless responding].

3.2.3. Item response times

Item response times are related to total questionnaire completion time, but they are focused on individual questionnaire items. Long item completion times may reflect poorly worded or difficult to understand questions. They may also reflect boredom or distraction at that point in the questionnaire. Exceedingly short item response times may reflect the fact that the respondent is ignoring the details of the question and responding randomly or to quickly move to the next item (see Gummer et al., 2021). Such "speeders" are defined variously as those who spend less than 1 or 2 s per item (Wood, Harms, Lowman, & DeSimone, 2017; Huang et al., 2012) or, more generously, less than 5.5 s (Curran, Kotrba, & Denison, 2010) [see Goldammer et al. (2020) for a recent detailed

analysis regarding response times in online questionnaires.].

3.2.4. Response accuracy

Response accuracy refers to the degree to which items on the questionnaire are answered correctly or completely. True item accuracy requires an expected or known response for its assessment. For most consumer opinion questionnaires this is not possible. However, response accuracy can be indirectly measured through the assessment of such suspicious behaviors as "straight-lining," defined by the number of times a respondent chooses the same response option, e.g., the LongString Index (Behrend, Sharek, Meade, & Wiebe, 2011; Johnson, 2005; Maniaci & Rogge, 2014b), overuse of the middle category on bi-directional scales, or random responding.

3.2.5. Response quality

Response quality refers to the degree to which the respondent fully and completely answers the question(s). One simple measure is the proportion of skipped items on a questionnaire or identification of unlikely or outlier responses (Maniaci & Rogge, 2014b). Response quality is especially important for open-ended questions that require the respondent to provide a text statement or answer. Here the completeness or depth of response can be gauged as a reflection of response quality. For multiple choice questions or questions requiring a numeric response, response quality may be reflected in item skipping. This is especially true for lengthy or involved questions, even though they may require a simple yes/no, multiple choice response or numeric entry.

4. Recommendations to assess and improve data quality in online questionnaires

Based on the above literature, the following recommendations can be made when obtaining data using online consumer questionnaires (Aim 3). These recommendations apply to all forms of research, both methodological and practical. While it is obvious that incorporating all of these recommendations in any one study may be impossible, given the practical limitations in many quarters of sensory and consumer research, to the extent that the data are collected with the intent to publish them in scientific journals, considerable effort should be placed on ensuring that the data are reliable, valid, and of the highest quality by adopting as many of these recommendations as possible.

In the sub-sections below, the findings and recommendations are organized in accordance with whether they relate to the survey and its administration, the respondent, or the survey taking environment. In addition, Table 2 summarizes these recommendations and provides guidance as to the most critical metrics and precautions that sensory and consumer researchers should undertake and report in published papers containing online questionnaire data.

Table 2
Examples of study information and data quality metrics that should be provided when reporting online questionnaire data in scientific journals.

1. Software platform used for questionnaire administration, database for participant selection, and any incentives to complete the questionnaire. If ISO certified vendor, so state.
2. A statement regarding the questionnaire-taking experience of the participants, e.g., average number of questionnaires per month; if available
3. Rates of participant completion, drop-outs, and post-hoc eliminations
4. Proportion of skipped or missing responses or statement that software prevented respondents from proceeding without completion of all previous questionnaire items
5. Task and/or survey completion times; if speeders eliminated, state criteria and percent eliminated
6. Protocols used to identify and eliminate non-human (bot) responses; include percent eliminated, if available
7. Post-hoc measures used to identify careless or inattentive participants, e.g., percent of straight-line or random responders
8. Ex-ante measures used to assess data quality, e.g., trap questions, instructed manipulation checks, etc.
9. Any measures used to assess participant engagement, topic interest or questionnaire enjoyment or satisfaction.
10. Any proactive or embedded manipulations to improve attention or engagement

4.1. Recommendations on survey design and administration

In general, using vendor panels is often a best practice for collecting online survey data, because of vendor capabilities to control access to the survey, eliminate bots, identify respondents, meet quotas, and control research-related elements of survey design and execution. In addition, many vendors are accredited as being in compliance with ISO Standard ISO 20252:2019, which specifies best practices in questionnaire research to ensure respondent identity and data quality through the use of data quality metrics (ISO, 2019). Crowd-sourced surveys may be demographically or psychographically skewed and should be used with caution, as should all forms of convenience sampling methods. Regardless of recruitment, it is important to assure respondents' interest in the survey topic. This can be accomplished through pre-screening questions or by using within-survey items that specifically address topic interest and general attitudes toward questionnaires and questionnaire taking.

Survey length is a critical factor in survey design. Long surveys induce boredom and inattentiveness. Any long survey should be divided into shorter segments varying in topic or response mode to avoid boredom, fatigue and drop-out. Important survey items should appear early in a survey, to garner greatest attention. Similarly, the longer a survey question or the greater the number of response alternatives, the poorer will be the data quality. Graphic analogue and pictorial response modes (e.g., emoji) are easier and more enjoyable for subjects to use than scales containing verbal labels, which are, in turn, better than unlabeled numeric category scales. However, the ease of scale usage must be considered within the context of the scale's desired psychometric properties and psychophysical measurement level.

To reach the widest available demographic population, Mobile Browser Surveys (MBS) should be used, i.e., those designed to operate on all devices (computers, tablets, smartphones). In addition, reasonable survey incentives will increase participation rates, but excessive incentives will attract respondents who may be only interested in obtaining the incentives.

4.2. Recommendations regarding respondents

The amount of experience with survey-taking by a respondent is a critical factor in data quality. Professional respondents must be avoided to minimize satisficing behavior and to ensure that the survey sample is similar to the target population in terms of product preferences and behaviors. However, data quality may be better among respondents with some degree of experience/literacy with survey data collection devices. Due to the widening presence of AI bots completing online surveys, CAPTCHA questions and other measures of determining response style should be employed in online surveys to eliminate contamination by data produced by AI bots.

Age of respondents does not have a major influence on data quality parameters, other than time to completion, which may be related to device familiarity. Gender appears to have no consistent effects in online survey data participation or quality. Educational level and cognitive ability have only minor effects on data quality and may, again, be more related to computer literacy and device competencies.

Respondent's Need for Cognition (NFC) and mindfulness may serve as measures and manipulations that link to survey attention/inattention and that can be used to improve online data quality in sensory and consumer questionnaires.

4.3. Recommendations regarding distractions, careless responding, and maladaptive behavior

Since it is impossible to control the environment in which online questionnaires are completed, careless responding and satisficing behaviors will be ubiquitous in online data. Therefore, precautions must be taken to identify and eliminate respondents exhibiting such behaviors through the use of both *post-hoc* and *ex-ante* measures.

Post-hoc measures of data quality should be employed by all sensory and consumer scientists collecting online data, to assess the validity of the data and to provide a basic index of data quality to readers of publications in which questionnaire data are reported. *Ex-ante* measures of data quality should also be used to further identify inattentive or maladaptive responding by consumers. A combination of *ex-ante* measures, including attention checks, trap questions, bogus and infrequency items, nonsense, and inconsistency checks should be used. Respondents who fail more than one of these items should have their data examined more closely for possible quality problems and for potential deletion from the survey. Further, since respondent engagement is a critical factor in questionnaire data quality, it should be quantified using direct measures of respondent engagement, involvement, interest and/or satisfaction with questionnaires or questionnaire topics. Efforts to improve respondent engagement in advance of survey administration through such techniques as gamification of the survey or embedded mindfulness training manipulations also should be considered.

5. Future research

We recommend that a concerted effort be undertaken by sensory and consumer researchers to assess and/or develop standardized *post hoc* and *ex ante* measures of data quality that are best suited to the study of food and other consumer products. These efforts should be focused on establishing scientifically valid criteria of good vs poor quality data, so that questionnaires and surveys in sensory and consumer studies can be compared on both an absolute and relative basis in terms of data quality. In addition, research should be directed toward the identification and/or development of proactive steps that can be taken to improve the quality of surveys and questionnaires, such as through the introduction of embedded pre-administration manipulations to improve respondent attention, cognitive focus and mindful responding. In all cases, the goal of such research and development efforts should be to produce the highest quality of online sensory and consumer data and to improve such practical outcomes as reduced data variability, higher effect sizes, and better questionnaire/product sensitivity and discriminability.

6. Funding statement

Financial support was received from two sources: 1) The New Zealand Institute for Plant and Food Research Limited and 2) The New Zealand Ministry for Business, Innovation & Employment.

CRediT authorship contribution statement

Sara R. Jaeger: Conceptualization, Writing – original draft, Writing – review & editing. **Armand V. Cardello:** Conceptualization, Writing – original draft, Writing – review & editing.

Declaration of Competing Interest

Given their roles as Editor at the time of submission, the authors were not involved in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another editor, as per the journal guidelines.

Acknowledgements

Staff at Plant & Food Research are thanked for help with editing.

References

Adigüzel, F., & Wedel, M. (2008). Split questionnaire design for massive surveys. *Journal of Marketing Research, 45*(5), 608–617.
Andreadis, I., & Kartsounidou, E. (2020). The impact of splitting a long online questionnaire on data quality. *Survey Research Methods, 14*(1), 31–42.
Andrews, F. M., & Herzog, A. R. (1986). The quality of survey data as related to age of respondent. *Journal of the American Statistical Association, 81*(394), 403–410.
Anduiza, E., & Galais, C. (2017). Answering without reading: IMCs and strong satisficing in online surveys. *International Journal of Public Opinion Research, 29*(3), 497–519.
Bangcuyo, R. G., Smith, K. J., Zumach, J. L., Pierce, A. M., Guttman, G. A., & Simons, C. T. (2015). The use of immersive technologies to improve consumer testing: The role of ecological validity, context and engagement in evaluating coffee. *Food Quality and Preference, 41*, 84–95.
Barge, S., & Gehlbach, H. (2012). Using the theory of satisficing to evaluate the quality of survey data. *Research in Higher Education, 53*(2), 182–200.
Baruch, Y. (1999). Response rate in academic studies: A comparative analysis. *Human Relations, 52*(4), 421–438.
Baumgartner, H., & Steenkamp, J. B. E. (2001). Response styles in marketing research: A cross-national investigation. *Journal of Marketing Research, 38*(2), 143–156.
Beach, D. A. (1989). Identifying the random responder. *The Journal of Psychology, 123*(1), 101–103.
Becker, R., Möser, S., & Glauser, D. (2019). Cash vs. vouchers vs. gifts in web surveys of a mature panel study—Main effects in a long-term incentives experiment across three panel waves. *Social Science Research, 81*, 221–234.
Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43*(3), 800–813.
Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2016). Can we turn shirkers into workers? *Journal of Experimental Social Psychology, 66*, 20–28.
Bi, J., & Kuesten, C. (2012). Intraclass correlation coefficient (ICC): A framework for monitoring and assessing performance of trained sensory panels and panelists. *Journal of Sensory Studies, 27*(5), 352–364.
Binswanger, J., Schunk, D., & Toepoel, V. (2013). Panel conditioning in difficult attitudinal questions. *Public Opinion Quarterly, 77*(3), 783–797.
Blasius, J., & Thiessen, V. (2001). The use of neutral responses in survey questions: An application of multiple correspondence analysis. *Journal of Official Statistics, 17*(3), 351–367.
Bohlmeijer, E., Ten Klooster, P. M., Fledderus, M., Veehof, M., & Baer, R. (2011). Psychometric properties of the five facet mindfulness questionnaire in depressed adults and development of a short form. *Assessment, 18*(3), 308–320.
Bosch, O. J., & Revilla, M. (2021). Using emojis in mobile web surveys for Millennials? A study in Spain and Mexico. *Quality & Quantity, 55*(1), 39–61.
Bowling, A. (2005). Mode of questionnaire administration can have serious effects on data quality. *Journal of public health, 27*(3), 281–291.
Bowling, N. A., Huang, J. L., Bragg, C. B., Khazon, S., Liu, M., & Blackmore, C. E. (2016). Who cares and who is careless? Insufficient effort responding as a reflection of respondent personality. *Journal of Personality and Social Psychology, 111*(2), 218.
Brown, K. W., & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(4), 822–848.
Brühlmann, F., Petralito, S., Aeschbach, L. F., & Opwis, K. (2020). The quality of data collected online: An investigation of careless responding in a crowdsourced sample. *Methods in Psychology, 2*, Article 100022.
Buskirk, T. D., & Andrus, C. (2012). Smart surveys for smart phones: Exploring various approaches for conducting online mobile surveys via smartphones. *Survey Practice, 5*(1), 1–11.
Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116–131.
Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*(3), 306–307.
Callegaro, M., Manfreda, K. L., & Vehovar, V. (2015). *Web Survey Methodology*. London: Sage.
Cannell, C. F., & Kahn, R. L. (1968). Interviewing. In G. Lindzey, & E. Aronson (Eds.), *The Handbook of Social Psychology* (pp. 526–595). Reading, MA: Addison-Wesley.
Cardello, A.V. and Schutz, H.G. (2010). Unpublished data.
Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior, 29*(6), 2156–2160.
Chang, L., & Krosnick, J. A. (2009). National surveys via RDD telephone interviewing versus the Internet: Comparing sample representativeness and response quality. *Public Opinion Quarterly, 73*(4), 641–678.
Chen, P. S. D. (2011). Finding quality responses: The problem of low-quality survey responses and its impact on accountability measures. *Research in Higher Education, 52*(7), 659–674.
Chong, M., Leung, A. K. Y., & Qing, V. L. Y. (2022). A cross-country investigation of social image motivation and acceptance of lab-grown meat in Singapore and the United States. *Appetite, 105990*.
Couper, M. P. (2000). Web surveys: A review of issues and approaches. *The Public Opinion Quarterly, 64*(4), 464–494.

Cox, E. P., III (1980). The optimal number of response alternatives for a scale: A review. *Journal of Marketing Research, 17*(4), 407–422.
Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66*, 4–19.
Curran, P. G., & Hauser, K. A. (2019). I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items. *Journal of Research in Personality, 82*, Article 103849.
Curran, P. G., Kotrba, L., & Denison, D. (2010). Careless responding in surveys: Applying traditional techniques to organizational settings. *Proceedings of the 25th Annual Conference of Society for Industrial and Organizational Psychology, Atlanta, GA*.
Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the index of consumer sentiment. *Public Opinion Quarterly, 64*(4), 413–428.
Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly, 69*(1), 87–98.
Deutskens, E., De Ruyter, K., Wetzels, M., & Oosterveld, P. (2004). Response rate and response quality of internet-based surveys: An experimental study. *Marketing Letters, 15*(1), 21–36.
de Winter, J. C., Kyriakidis, M., Dodou, D., & Happee, R. (2015). Using CrowdFlower to study the relationship between self-reported violations and traffic accidents. *Procedia Manufacturing, 3*, 2518–2525.
De Wulf, K., & Berteloot, S. (2007). Duplication and multi-source online panel recruitment: Real quality differences or idle rumours. In *Proceedings of ESOMAR World Research Conference on Panel Research* (pp. 49–62).
Ding, M., Grewal, R., & Liechty, J. (2005). Incentive-aligned conjoint analysis. *Journal of Marketing Research, 42*(1), 67–82.
Douglas, K. M., & McGarty, C. (2001). Identifiability and self-presentation: Computer-mediated communication and intergroup interaction. *British Journal of Social Psychology, 40*(3), 399–416.
Downes-LeGuin, T. D. L., Baker, R., Mechling, J., & Ruyle, E. (2012). Myths and realities of respondent engagement in online surveys. *International Journal of Market Research, 54*(5), 613–633.
Dupuis, M., Meier, E., Gholam-Rezaee, M., Gmel, G., Strippoli, M. P. F., & Renaud, O. (2020). Detecting computer-generated random responding in online questionnaires: An extension of Dupuis, Meier & Cuneo (2019) on dichotomous data. *Personality and Individual Differences, 157*, Article 109812.
Eisenberger, R., & Cameron, J. (1996). Detrimental effects of reward: Reality or myth? *American Psychologist, 51*(11), 1153–1166.
Evans, J. R., & Mathur, A. (2005). The value of online surveys. *Internet Research, 15*, 195–219.
Evans, J. R., & Mathur, A. (2018). The value of online surveys: A look back and a look ahead. *Internet Research, 28*, 854–887.
Fan, W., & Yan, Z. (2010). Factors affecting response rates of the web survey: A systematic review. *Computers in Human Behavior, 26*(2), 132–139.
Frede, S., Markowitz, L., & Coffery, S. (2006). Heavier Responders in Online Survey Research. *White Paper, Ipsos, NPD Group, and TNS.*
Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25–42.
Fricker, S., Creech, B., Davis, J., Gonzalez, J., Tan, L., & To, N. (2012). Does length really matter? Exploring the effects of a shorter interview on data quality, nonresponse, and respondent burden. *Federal Committee on Statistical Methodology 2012 Research Conference, Washington, DC, USA*.
Fricker, S., Galesic, M., Tourangeau, R., & Yan, T. (2005). An experimental comparison of web and telephone surveys. *Public Opinion Quarterly, 69*(3), 370–392.
Gadiraju, U., Kawase, R., Dietze, S., & Demartini, G. (2015). Understanding malicious behavior in crowdsourcing platforms: The case of online surveys. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, Seoul, Republic of Korea* (pp. 1631-1640).
Galesic, M. (2005). Anketna istraživanja putem interneta: Mogući izvori pogrešaka. *Društvena istraživanja-Časopis za opća društvena pitanja, 14*(75+76), 297–320.
Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics, 22*(2), 313–328.
Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*(2), 349–360.
Ganassali, S. (2008). The influence of the design of web survey questionnaires on the quality of responses. *Survey Research Methods, 2*(1), 21–32.
Gao, Z., House, L. A., & Xie, J. (2016). Online survey data quality and its implication for willingness-to-pay: A cross-country comparison. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie, 64*(2), 199–221.
Gao, Z., House, L., & Bi, X. (2016). Impact of satisficing behavior in online surveys on consumer preference and welfare estimates. *Food Policy, 64*, 26–36.
Garland, P., Santus, D., & Uppal, R. (2009). Survey lockouts: Are we too cautious? *The Panel Conference of the Council of American Survey Research Organizations, New Orleans, USA*.
Gittelman, S., & Trimarchi, E. (2010). Online Research… And All That Jazz! The practical adaptation of old tunes to make new music. *ESOMAR Online Research*. https://www.mktginc.com/pdf/berlin.pdf.
Goldammer, P., Annen, H., Stöckli, P. L., & Jonas, K. (2020). Careless responding in questionnaire measures: Detection, impact, and remedies. *The Leadership Quarterly, 31*(4), Article 101384.
Göritz, A. S. (2006). Incentives in web studies: Methodological issues and a review. *International Journal of Internet Science, 1*(1), 58–70.
Goyder, J., Warriner, K., & Miller, S. (2002). Evaluating socio-economic status (SES) bias in survey nonresponse. *Journal of Official Statistics, 18*(1), 1–12.
Greszki, R., Meyer, M., & Schoen, H. (2014). The impact of speeding on data quality in nonprobability and freshly recruited probability-based online panels. *Online Panel Research: A Data Quality Perspective*, 238-262.
Greszki, R., Meyer, M., & Schoen, H. (2015). Exploring the effects of removing "too fast" responses and respondents from web surveys. *Public Opinion Quarterly, 79*(2), 471–503.
Groves, R. M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly, 68*(1), 2–31.
Guidry, K. R. (2012). Response quality and demographic characteristics of respondents using a mobile device on a web-based survey. *Paper presented at the Annual Meeting of the American Association for Public Opinion Research, Orlando, FL* (pp. 17-20).
Gummer, T., & Roßmann, J. (2013). Good questions, bad questions? A post-survey evaluation strategy based on item nonresponse. *Survey Methods: Insights from the Field*, 1-10.
Gummer, T., Roßmann, J., & Silber, H. (2021). Using instructed response items as attention checks in web surveys: Properties and implementation. *Sociological Methods & Research, 50*(1), 238–264.
Hannum, M. E., Forzley, S., Popper, R., & Simons, C. T. (2020). Further validation of the engagement questionnaire (EQ): Do immersive technologies actually increase consumer engagement during wine evaluations? *Food Quality and Preference, 85*, Article 103966.
Hannum, M. E., & Simons, C. T. (2020). Development of the engagement questionnaire (EQ): A tool to measure panelist engagement during sensory and consumer evaluations. *Food Quality and Preference, 81*, Article 103840.
Harms, J., Biegler, S., Wimmer, C., Kappel, K., & Grechenig, T. (2015). Gamification of online surveys: Design process, case study, and evaluation. In *Proceedings of IFIP Conference on Human-Computer Interaction – INTERACT 2015* (pp. 219–236).
Harms, C., Jackel, L., & Montag, C. (2017). Reliability and completion speed in online questionnaires under consideration of personality. *Personality and Individual Differences, 111*, 281–290.
Harms, J., Wimmer, C., Kappel, K., & Grechenig, T. (2014). Gamification of online surveys: Conceptual foundations and a design process based on the MDA framework. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational. Helsinki, Finland* (pp. 565–568).
Hathaway, D., & Simons, C. T. (2017). The impact of multiple immersion levels on data quality and panelist engagement for the evaluation of cookies under a preparation-based scenario. *Food Quality and Preference, 57*, 114–125.
Hauser, D. J., & Schwarz, N. (2015). It's a trap! Instructional manipulation checks prompt systematic thinking on "tricky" tasks. *Sage Open, 5*(2), 2158244015584617.
Hays, R. D., Liu, H., & Kapteyn, A. (2015). Use of Internet panels to conduct surveys. *Behavior Research Methods, 47*(3), 685–690.
Heyman, J., & Ariely, D. (2004). Effort for payment: A tale of two markets. *Psychological science, 15*(11), 787–793.
Hillygus, D. S., Jackson, N., & Young, M. (2014). Professional respondents in non-probability online panels. *Online panel research: A data quality perspective, 1*, 219–237.
Holbrook, A. L., Anand, S., Johnson, T. P., Cho, Y. I., Shavitt, S., Chávez, N., & Weiner, S. (2014). Response heaping in interviewer-administered surveys: Is it really a form of satisficing? *Public Opinion Quarterly, 78*(3), 591–633.
Honda, N., & Motokawa, A. (2005). Can internet surveys be used for social surveys?: Results of an experimental study. *The Japan Institute for Labour Policy and Training Research Report, 15*, 1–26.
Huang, J. L., Bowling, N. A., Liu, M., & Li, Y. (2014). Detecting insufficient effort responding with an infrequency scale: Evaluating validity and participant reactions. *Journal of Business and Psychology, 30*(2), 299–311.
Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology, 27*(1), 99–114.
Huang, J. L., Liu, M., & Bowling, N. A. (2015). Insufficient effort responding: Examining an insidious confound in survey data. *Journal of Applied Psychology, 100*(3), 828–845.
Hull, R., & Reid, J. (2003). Designing engaging experiences with children and artists. In M. A. Blythe, A. F. Monk, K. Overbeeke, & P. C. Wright (Eds.), *Funology: From Usability to Enjoyment* (pp. 179-187). Dordrecht, The Netherlands: Kluwer.
International Standards Organization. (2019). *ISO20252:2019 – Market, opinion and social research, including insights and data analytics – vocabulary and service requirement*. Geneva: International Organization for Standardization.
Jaeger, S. R., & Cardello, A. V. (2009). Direct and indirect hedonic scaling methods: A comparison of the labeled affective magnitude (LAM) scale and best–worst scaling. *Food Quality and Preference, 20*(3), 249–258.
Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality, 39*(1), 103–129.
Jones, M. S., House, L. A., & Gao, Z. (2015). Respondent screening and revealed preference axioms: Testing quarantining methods for enhanced data quality in web panel surveys. *Public Opinion Quarterly, 79*(3), 687–709.
Keusch, F., & Zhang, C. (2017). A review of issues in gamified surveys. *Social Science Computer Review, 35*(2), 147–166.
Knapton, K., & Garlick, R. (2007). Catch me if you can. *Quirks, November*, 58-63.
Kreuter, F., Presser, S., & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and web surveys the effects of mode and question sensitivity. *Public Opinion Quarterly, 72*(5), 847–865.
Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213–236.
Krosnick, J. A., & Alwin, D. F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly, 51*(2), 201–219.

Krosnick, J. A., & Berent, M. K. (1993). Comparisons of party identification and policy preferences: The impact of survey question format. *American Journal of Political Science*, 941–964.
Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. *Survey Measurement and Process Quality*, 141–164.
Kruse, Y., Callegaro, M., Dennis, J. M., Subias, S., Lawrence, M., DiSogra, C., & Tompson, T. (2009). *Panel conditioning and attrition in the AP-Yahoo! news election panel study*. Hollywood, Florida: American Association for Public Opinion Research (AAPOR).
Kutschar, P., Weichbold, M., & Osterbrink, J. (2019). Effects of age and cognitive function on data quality of standardized surveys in nursing home populations. *BMC Geriatrics, 19*(1), 1–10.
Laurel, B. (1993). Engagement: The First-Person Imperative. *Chapter Four-Dramatic Techniques for Orchestrating Human Response*. In D. J. DeRose (Ed.), *Computers as Theatre* (pp. 111–125). Reading, MA: Addison-Wesley Publishing Company.
Lawless, H. T., Sinopoli, D., & Chapman, K. W. (2010). A comparison of the labeled affective magnitude scale and the 9-point hedonic scale and examination of categorical behavior. *Journal of Sensory Studies, 25*, 54–66.
Lee, J., & Paek, I. (2014). In search of the optimal number of response categories in a rating scale. *Journal of Psychoeducational Assessment, 32*(7), 663–673.
Lim, J. (2011). Hedonic scaling: A review of methods and theory. *Food Quality and Preference, 22*(8), 733–747.
Lindhjem, H., & Navrud, S. (2011). Are Internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics, 70*(9), 1628–1637.
Liu, M., & Wronski, L. (2018). Examining completion rates in web surveys via over 25,000 real-world surveys. *Social Science Computer Review, 36*(1), 116–124.
Louviere, J. J., Flynn, T. N., & Marley, A. A. J. (2015). *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge: Cambridge University Press.
Lund, C. M., Jones, V. S., & Spanitz, S. (2009). Effects and influences of motivation on trained panelists. *Food Quality and Preference, 20*(4), 295–303.
Lyberg, L. (2012). Survey quality. *Survey. Methodology, 38*(2), 107–130.
MacLean, K. A., Ferrer, E., Aichele, S. R., Bridwell, D. A., Zanesco, A. P., Jacobs, T. L., King, B. G., Rosenberg, E. L., Sahdra, B. K., Shaver, P. R., Wallace, B. A., Mangun, G. R., & Saron, C. D. (2010). Intensive meditation training improves perceptual discrimination and sustained attention. *Psychological Science, 21*(6), 829–839.
Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly, 72*(5), 914–934.
Maloney, P., Grawitch, M. J., & Barber, L. K. (2011). Strategic item selection to reduce survey length: Reduction in validity? *Consulting Psychology Journal: Practice and Research, 63*(3), 162–175.
Mancosu, M., Ladini, R., & Vezzoni, C. (2019). 'Short is Better'. Evaluating the Attentiveness of Online Respondents Through Screener Questions in a Real Survey Environment. *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 141*(1), 30–45.
Maniaci, M. R., & Rogge, R. D. (2014a). Conducting research on the Internet. In H. T. Reis, & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (pp. 443–470). Cambridge: Cambridge University Press.
Maniaci, M. R., & Rogge, R. D. (2014b). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality, 48*, 61–83.
Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*(4), 810–819.
Mason, W., & Watts, D. J. (2009). Financial incentives and the" performance of crowds". In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 77–85).
Matell, M. S., & Jacoby, J. (1971). Is there an optimal number of alternatives for Likert scale items? Study I: Reliability and validity. *Educational and Psychological Measurement, 31*(3), 657–674.
Mavletova, A. (2013). Data quality in PC and mobile web surveys. *Social Science Computer Review, 31*(6), 725–743.
McEwan, J. A., Hunter, E. A., van Gemert, L. J., & Lea, P. (2002). Proficiency testing for sensory profile panels: Measuring panel performance. *Food Quality and Preference, 13*(3), 181–190.
Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*(3), 437–455.
Meiselman, H. L., Waterman, D., & Symington, L. E. (1974). *Armed forces food preferences*. Natick, MA: Army Natick Development Center.
Menold, N. (2020). Rating-scale labeling in online surveys: An experimental comparison of verbal and numeric rating scales with respect to measurement quality and respondents' cognitive processes. *Sociological Methods & Research, 49*(1), 79–107.
Menold, N., Kaczmirek, L., Lenzner, T., & Neusar, A. (2014). How do respondents attend to verbal labels in rating scales? *Field Methods, 26*(1), 21–39.
Miller, J. (2007). *Burke panel quality research and development summary. White Paper: Advertising Research Foundation Online Research Quality Council*. Cincinnati: Burke and Company.
Moore, D. L., & Tarnai, J. (2002). Evaluating nonresponse error in mail surveys. In R. M. Groves, D. A. Dillman, J. L. Eltinge, & R. J. A. Little (Eds.), *Survey Nonresponse* (pp. 197–211). New York: John Wiley & Sons.
Morii, M., Sakagami, T., Masuda, S., Okubo, S., & Tamari, Y. (2017). How does response bias emerge in lengthy sequential preference judgments? *Behaviormetrika, 44*(2), 575–591.
Moss, A. & Litman, L. (2022). The CloudResearch Guide to Data Quality, Part 1: How to Define and Assess Data Quality in Online Research. https://www.cloudresearch.com/ resources /guides/ultimate-guide-to-survey-data-quality/guide-data-quality-what-is-data-quality-why-important/. Accessed 25 June 2022.
Mrazek, M. D., Franklin, M. S., Phillips, D. T., Baird, B., & Schooler, J. W. (2013). Mindfulness training improves working memory capacity and GRE performance while reducing mind wandering. *Psychological Science, 24*(5), 776–781.
Nancarrow, C., & Cartwright, T. (2007). Online access panels and tracking research: The conditioning issue. *International Journal of Market Research, 49*(5), 573–594.
Nielsen, J. S. (2011). Use of the Internet for willingness-to-pay surveys: A comparison of face-to-face and web-based interviews. *Resource and Energy Economics, 33*(1), 119–129.
Niemi, R. G., Carmines, E. G., & McIver, J. P. (1986). The impact of scale length on reliability and validity. *Quality and Quantity, 20*(4), 371–376.
Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use? *Journal of Research in Personality, 63*, 1–11.
Nikolaus, C. J., Ellison, B., & Nickols-Richardson, S. M. (2020). Food Insecurity among college students differs by questionnaire modality: An exploratory study. *American Journal of Health Behavior, 44*(1), 82–89.
O'Brien, H. L., & Toms, E. G. (2005). *Engagement as process in human-computer interactions*. 42(1): American Society for Information Science and Technology.
O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology, 59*(6), 938–955.
O'Brien, H. L., & Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. *Journal of the American Society for Information Science and Technology, 61*(1), 50–69.
Olsen, S. B. (2009). Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. *Environmental and Resource Economics, 44*(4), 591–610.
Ono, H., & Zavodny, M. (2003). Gender and the Internet. *Social Science Quarterly, 84*(1), 111–121.
Ono, H., & Zavodny, M. (2016). Internet and gender. In N. Naples, R. C. Hoogland, & M. Wickramasinghe (Eds.), *The Wiley Blackwell Encyclopedia of Gender and Sexuality Studies*. Oxford: Wiley-Blackwell.
Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867–872.
Peytchev, A., & Tourangeau, R. (2005). *How questionnaire layout induces measurement error*. Miami Beach, FL: Annual Meeting of the American Association for Public Opinion Research.
Pineau, V., Nukulkij, P., & Tang, X. (2005). Assessing panel bias in the Knowledge Networks panel: Updated results from 2005 research. In American Statistical Association (Ed.), JSM Proceedings [CD-ROM] (pp. 3480-3486). Alexandria, VA: AMSTAT.
Pliner, P., & Hobden, K. (1992). Development of a scale to measure the trait of food neophobia in humans. *Appetite, 19*, 105–120.
Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.
Polonsky, M. J., & Vocino, A. (2010). Survey completion speed of online panelists: the role of demographics and experience. *Proceedings of the 2010 Australian and New Zealand Marketing Academy Conference, Christchurch, New Zealand*.
Porter, S. R., & Whitcomb, M. E. (2003). The impact of lottery incentives on student survey response rates. *Research in Higher Education, 44*(4), 389–407.
Revilla, M., & Ochoa, C. (2017). Ideal and maximum length for a web survey. *International Journal of Market Research, 59*(5), 557–565.
Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2014). Do online access panels really need to allow and adapt surveys to mobile devices? RECSM Working Paper Number 41. Barcelona: Universitat Pompeu Fabra. Retrieved from www.upf.edu/survey/_pdf/RECSM_wp041.pdf (Accessed 11 June 2021)3.
Revilla, M., Toninelli, D., Ochoa, C., & Loewe, G. (2016). Do online access panels need to adapt surveys for mobile devices? *Internet Research, 26*(5), 1209–1227.
Robinson, M. A., & Boies, K. (2021). On the quest for quality self-report data: HEXACO and indicators of careless responding. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 53*(3), 377–380.
Rogelberg, S. G., Fisher, G. G., Maynard, D. C., Hakel, M. D., & Horvath, M. (2001). Attitudes toward surveys: Development of a measure and its relationship to respondent behavior. *Organizational Research Methods, 4*(1), 3–25.
Rogelberg, S. G., Luong, A., Sederburg, M. E., & Cristol, D. S. (2000). Employee attitude surveys: Examining the attitudes of noncompliant employees. *Journal of Applied Psychology, 85*(2), 284–293.
Roßmann, J. (2010). Data quality in web surveys of the German longitudinal election study 2009. The *3rd ECPR Graduate Conference, Dublin, Ireland* (Vol. 30).
Roßmann, J., Gummer, T., & Silber, H. (2018). Mitigating satisficing in cognitively demanding grid questions: Evidence from two web-based experiments. *Journal of Survey Statistics and Methodology, 6*(3), 376–400.
Roster, C. A., Albaum, G., & Smith, S. M. (2017). Effect of topic sensitivity on online survey panelists' motivation and data quality. *Journal of Marketing Theory and Practice, 25*(1), 1–16.
Sandorf, E. D., Persson, L., & Broberg, T. (2020). Using an integrated choice and latent variable model to understand the impact of "professional" respondents in a stated preference survey. *Resource and Energy Economics, 61*, Article 101178.
Saris, W. E., & Gallhofer, I. (2007). Estimation of the effects of measurement characteristics on the quality of survey questions. *Survey Research Methods, 1*(1), 29–43.
Saxon, D., Garratt, D., Gilroy, P., & Cairns, C. (2003). Collecting data in the information age: Exploring Web-based survey methods in educational research. *Research in Education, 69*(1), 51–66.

Schifferstein, H. N., Lemke, M., & de Boer, A. (2022). An exploratory study using graphic design to communicate consumer benefits on food packaging. *Food Quality and Preference, 97*, Article 104458.
Schonlau, M., & Toepoel, V. (2015). Straightlining in Web survey panels over time. *Survey Research Methods, 9*(2), 125–137.
Seelye, A., Mattek, N., Sharma, N., Riley, T., Austin, J., Wild, K., Dodge, H. H., Lore, E., & Kaye, J. (2018). Weekly observations of online survey metadata obtained through home computer use allow for detection of changes in everyday cognition before transition to mild cognitive impairment. *Alzheimer's & Dementia, 14*(2), 187–194.
Shamon, H., & Berning, C. (2020). Attention Check Items and Instructions in Online Surveys with Incentivized and Non-Incentivized Samples: Boon or Bane for Data Quality? *Survey Research Methods, 14*(1), 55–77.
Simon, H. A. (1957). *Models of Man; Social and Rational*. New York: Wiley.
Simon, H. A., Stedry, A. C., Lindzey, G., & Aronson, E. (1968). Psychology and Economics. In G. Lindzey, & E. Aronson (Eds.), *The Handbook of Social Psychology* (pp. 269–314). Reading, MA: Addison-Wesley.
Simone, M. (2019) First Opinion: Bots started sabotaging my online research. I fought back. STAT. Retrieved from https://www.statnews.com/2019/11/21/bots-started-sabotaging-my-online-research-i-fought-back/ Accessed 11 June 2021).
Singer, E., Van Hoewyk, J., & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly, 64*(2), 171–188.
Sipos, L., Nyitrai, Á., Hitka, G., Friedrich, L. F., & Kókai, Z. (2021). Sensory Panel Performance Evaluation—Comprehensive Review of Practical Approaches. *Applied Sciences, 11*(24), 11977.
Sjoberg, G. (1954). A questionnaire on questionnaires. *Public Opinion Quarterly, 18*(4), 423–427.
Smith, G. (2008). Does gender influence online survey participation? A record-linkage analysis of university faculty online survey response behavior. *ERIC Document Reproduction Service No. ED 501717*.
Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment, 12*(1), 102–111.
Smith, R., & Brown, H. H. (2006). *Panel and data quality: comparing metrics and assessing claims*. Barcelona, Spain: Panel Research.
Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research, 69*(8), 3139–3148.
Stähli, E., & Joye, D. (2016). Incentives as a possible measure to increase response rates. *The SAGE Handbook of Survey Methodology*, 425–440.
Steptoe, A., Pollard, T. M., & Wardle, J. (1995). Development of a measure of the motives underlying the selection of food: The food choice questionnaire. *Appetite, 25*(3), 267–284.
Storozuk, A., Ashley, M., Delage, V., & Maloney, E. A. (2020). Got bots? Practical recommendations to protect online survey data from bot attacks. *The Quantitative Methods for Psychology, 16*(5), 472–481.
Subar, A. F., Ziegler, R. G., Thompson, F. E., Johnson, C. C., Weissfeld, J. L., Reding, D., Kavounis, K. H., & Hayes, R. B. (2001). Is shorter always better? Relative importance of questionnaire length and cognitive ease on response rates and data quality for two dietary questionnaires. *American Journal of Epidemiology, 153*(4), 404–409.
Toepoel, V. (2016). Developing the Survey: Questions and Answers. *Doing Surveys Online, 55*, 19–38.
Toepoel, V., Das, M., & Van Soest, A. (2008). Effects of design in web surveys: Comparing trained and fresh respondents. *Public Opinion Quarterly, 72*(5), 985–1007.
Toepoel, V., & Dillman, D. A. (2011). Words, numbers, and visual heuristics in web surveys: Is there a hierarchy of importance? *Social Science Computer Review, 29*(2), 193–207.
Toepoel, V., & Funke, F. (2018). Sliders, visual analogue scales, or buttons: Influence of formats and scales in mobile and desktop surveys. *Mathematical Population Studies, 25*(2), 112–122.
Toepoel, V., & Ludtig, P. (2015). Online surveys are mixed-device surveys. Issues associated with the use of different (mobile) devices in web surveys. *Methods, Data, Analyses, 9*(2), 155–162.
Tomic, O., Forde, C., Delahunty, C., & Næs, T. (2013). Performance indices in descriptive sensory analysis–A complimentary screening tool for assessor and panel performance. *Food Quality and Preference, 28*(1), 122–133.
Tonsor, G. T., Lusk, J. L., & Schroeder, T. C. (2022). Market potential of new plant-based protein alternatives: Insights from four US consumer experiments. *Applied Economic Perspectives and Policy*, 1–18.
Tourangeau, R. (1984). *Cognitive science and survey methods. Cognitive aspects of survey methodology: Building a bridge between disciplines* (pp. 73–100). Washington, DC: National Academy Press.
Tourangeau, R., Couper, M. P., & Conrad, F. (2007). Color, labels, and interpretive heuristics for response scales. *Public Opinion Quarterly, 71*(1), 91–112.
Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. United Kingdom: Cambridge University Press.
Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly, 60*(2), 275–304.
Tourangeau, R., Sun, H., Yan, T., Maitland, A., Rivero, G., & Williams, D. (2018). Web surveys by smartphones and tablets: Effects on data quality. *Social Science Computer Review, 36*(5), 542–556.
Triantoro, T., Gopal, R., Benbunan-Fich, R., & Lang, G. (2019). Would you like to play? A comparison of a gamified survey with a traditional online survey method. *International Journal of Information Management, 49*, 242–252.
Triantoro, T., Gopal, R., Benbunan-Fich, R., & Lang, G. (2020). Personality and games: Enhancing online surveys through gamification. *Information Technology and Management, 21*(3), 169–178.
Vannette, D. L. (2016). *Testing the effects of different types of attention interventions on data quality in web surveys. Experimental evidence from a 14-country study*. Austin, TX: American Association for Public Opinion Research.
Vaughn, J. (2022) Data quality. Tech Target, https://www.techtarget.com/searchdatamanagement/definition/data-quality, accessed April 30, 2022.
Walker, R., Pettit, R., & Rubinson, J. (2009). The foundations of quality initiative: A five-part immersion into the quality of online research. *Journal of Advertising Research, 49*(4), 464–485.
Wansink, B., & Sudman, S. (2002). Predicting the future of consumer panels. *Journal of Database Marketing & Customer Strategy Management, 9*(4), 301–311.
Weijters, B., & Baumgartner, H. (2012). Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research, 49*(5), 737–747.
Weng, L. J. (2004). Impact of the number of response categories and anchor labels on coefficient alpha and test-retest reliability. *Educational and Psychological measurement, 64*(6), 956–972.
Willems, P., Van Ossenbruggen, R., & Vonk, T. (2006). The effects of panel recruitment and management on research results: A study across 19 online panels. *Paper presented at ESOMAR Panel Research Conference, Barcelona, Spain*.
Windschitl, P. D., & Wells, G. L. (1996). Measuring psychological uncertainty: Verbal versus numeric methods. *Journal of Experimental Psychology: Applied, 2*(4), 343.
Wong, N., Rindfleisch, A., & Burroughs, J. E. (2003). Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale. *Journal of Consumer Research, 30*(1), 72–91.
Wood, D., Harms, P. D., Lowman, G. H., & DeSimone, J. A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychological and Personality Science, 8*(4), 454–464.
Wright, K. B. (2005). Researching Internet-based populations: Advantages and disadvantages of online survey research, online questionnaire authoring software packages, and web survey services. *Journal of Computer-Mediated. Communication, 10*(3), JCMC1034.
Yan, T., & Tourangeau, R. (2008). Fast times and easy questions: The effects of age, experience and question complexity on web survey response times. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 22*(1), 51–68.
Zhang, B., & Gearhart, S. (2020). Collecting Online Survey Data: A Comparison of Data Quality among a Commercial Panel & MTurk. *Survey Practice, 13*(1), 1–10.
Zhang, C., & Conrad, F. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straight-lining. *Survey Research Methods, 8*(2), 127–135.