

Association for Information Systems  
**AIS Electronic Library (AISeL)**

---

ICIS 2025 Proceedings

International Conference on Information  
Systems (ICIS)

---

December 2025

## **Impact of Explicit and Hidden Survey Framing on Paid Online Survey Panel Participant Screening Rates, Attentiveness, and Data Quality: An Empirical Examination**

Henri Jalo  
*Tampere University*, [henri.jalo@tuni.fi](mailto:henri.jalo@tuni.fi)

Markus Makkonen  
*Tampere University*, [markus.makkonen@tuni.fi](mailto:markus.makkonen@tuni.fi)

Henri Pirkkalainen  
*Tampere University*, [henri.pirkkalainen@tuni.fi](mailto:henri.pirkkalainen@tuni.fi)

Marko Seppänen  
*Tampere University*, [marko.seppanen@tut.fi](mailto:marko.seppanen@tut.fi)

Follow this and additional works at: <https://aiselaisnet.org/icis2025>

---

### **Recommended Citation**

Jalo, Henri; Makkonen, Markus; Pirkkalainen, Henri; and Seppänen, Marko, "Impact of Explicit and Hidden Survey Framing on Paid Online Survey Panel Participant Screening Rates, Attentiveness, and Data Quality: An Empirical Examination" (2025). *ICIS 2025 Proceedings*. 4.  
[https://aiselaisnet.org/icis2025/is\\_researchmethods/is\\_researchmethods/4](https://aiselaisnet.org/icis2025/is_researchmethods/is_researchmethods/4)

This material is brought to you by the International Conference on Information Systems (ICIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in ICIS 2025 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# **Impact of Explicit and Hidden Survey Framing on Paid Online Survey Panel Participant Screening Rates, Attentiveness, and Data Quality: An Empirical Examination**

*Completed Research Paper*

**Henri Jalo**

Tampere University

P.O. Box 553, 33104, Tampere, Finland  
henri.jalo@tuni.fi

**Markus Makkonen**

Tampere University

P.O. Box 553, 33104, Tampere, Finland  
markus.makkonen@tuni.fi

**Henri Pirkkalainen**

Tampere University

P.O. Box 553, 33104, Tampere, Finland  
henri.pirkkalainen@tuni.fi

**Marko Seppänen**

Tampere University

P.O. Box 300, 33104, Tampere,  
Finland  
marko.seppanen@tuni.fi

## **Abstract**

*Researchers increasingly use paid online panels to collect survey data. However, financial incentives may lead to inappropriate participant self-selection, compromising data quality and relevance. Despite these concerns, the fundamental question of how participants are screened into surveys has received surprisingly scant attention. This study examines the effect of alternative survey framing approaches on screen-out rates, attentiveness, and data quality. Two samples were collected via Prolific: one with participant opt-in based on explicit topic disclosure ( $n = 1,035$ ), and the other using hidden framing with screening criteria concealed from respondents ( $n = 695$ ). Analysis revealed a substantial difference in screen-out rates (explicit: 11.87%, hidden: 78.33%), but no significant difference in data quality based on reliability or convergent and discriminant validity. However, common method variance was unevenly distributed in the explicit framing sample and evenly distributed in the hidden framing sample. Based on these findings, we offer practical recommendations for survey researchers.*

**Keywords:** survey, panel, methodology, self-selection, screening, filtering, honesty, integrity, attentiveness, data quality, common method variance.

## **Introduction**

Online survey panels (e.g., Amazon MTurk and Prolific) have become indispensable tools for information systems (IS) researchers (Casler et al., 2013; Kan & Drumme, 2018). These panels provide a wide range of respondents fitting a multitude of research contexts, making survey data collection swift and convenient compared to collecting data from professional networks or organizations (Hays et al., 2015) or from students, where the problem of generalizability to other contexts is a prevalent concern (Compeau et al., 2012). However, because panel participants are paid for their responses, this raises a critical concern over inappropriate self-selection (Chandler & Paolacci, 2017). Financially motivated respondents may attempt to complete as many surveys as possible to maximize their rewards, even when the survey context is

irrelevant to them (Chandler & Paolacci, 2017; Hays et al., 2015). Respondents without domain experience are likely to provide data that is of lower quality and relevance (Hays et al., 2015), biasing the inferences that are made from the whole data set (Pickering & Blaszczyński, 2021). Respondents posing as members of the target population to gain access to paid surveys is reported to be extremely common in crowdsourced survey studies (Sharpe Wessling et al., 2017), and some methodologists even consider such impostor participants to be the greatest threat to data quality (Chandler & Paolacci, 2017).

The reliability and integrity of inferences drawn from survey panel data are critical, as they can shape the direction of research and influence the investment decisions of companies that rely on an accurate understanding of end-user perspectives. Although some studies suggest that data collected from online panels are of comparable quality to those collected from laboratory settings or via social media (Casler et al., 2013), concerns remain. Moreover, in most research articles, the recruitment of survey participants is reported opaquely or not at all (Stefkovics et al., 2024). It is often unclear whether anyone with access to the survey link could have responded regardless of eligibility, or whether participants were deftly screened in based on criteria that were undisclosed to them beforehand. This is a critical issue, as responses are likely to be uninformative if participants lack genuine experience (Chandler & Paolacci, 2017).

With the proliferation of online survey panels and significant increases in pools of panel respondents (Peer et al., 2022), the distance between researchers and survey respondents has widened. Researchers now rely on internal screening mechanisms offered by the panel platforms to select relevant respondents. Recently, some online panels<sup>1</sup> have enabled the use of self-developed custom screening questions within surveys, beyond the more limited internal screeners provided by the platform. Platforms such as Prolific<sup>2</sup> recommend designing screening questions to be vague and obfuscated, making it difficult for survey participants to infer the actual survey topic and thereby reducing dishonest self-selection. This indicates that panel platforms also recognize the severity of the issue. However, it remains unclear whether data collected from respondents selected through concealed screening criteria is of higher quality than data collected from self-selected respondents who are explicitly informed of the survey theme.

Due to in-study screening features becoming available only recently, there is a critical gap in our methodological understanding on how online panel recruitment should be handled (Sharpe Wessling et al., 2017). Specifically, it is unclear whether the aim of the survey should be openly disclosed at the recruitment phase or concealed until respondents have been screened in based on hidden criteria, and what impact these different framing approaches have on data quality metrics. *The objective of this research is to assess the effect of two alternative survey framing approaches on survey participant screening rates, attentiveness, and data quality.* To assess the outcome differences of the framing approaches, this study examines whether any differences emerge between data sets collected through (1) user-led opt in based on an upfront description of the survey topic and (2) hidden framing using in-survey screening, where the broad topic of the survey is initially described, and respondents are subsequently screened in based on concealed eligibility criteria. This study contributes to IS methodology literature by comparing two different samples collected from Prolific using explicit survey framing ( $n = 1,035$ ) and hidden framing ( $n = 695$ ). The substantial survey examined the impact of technostress, organizational factors, and technological aspects on adoption resistance and attitudes toward adopting electric work vehicles (EWVs) in non-road contexts.

The results show a significant difference in screen-out rates: only 11.87% of respondents screened themselves out in the explicit framing condition, compared to 78.33% who were screened out based on concealed criteria in the hidden framing. The explicit framing sample completed the survey faster, but there was no difference in passing attention trap questions, suggesting high attentiveness. Self-reported measures of participant engagement were also statistically equivalent. No substantial differences were identified in construct reliability, convergent validity, or discriminant validity, as measurement invariance between the samples was established, indicating high surface-level data quality. However, differences in common method variance (CMV) were identified between the samples: the explicit framing sample exhibited unevenly distributed CMV concentrated on positively valenced items, whereas CMV in the hidden framing

---

<sup>1</sup> Prolific's announcement on in-study screener feature <https://www.prolific.com/resources/what-s-new-custom-screening-questions-smoother-study-setup-and-more>

<sup>2</sup> Prolific's recommendations for screeners <https://researcher-help.prolific.com/en/article/6bad1f>

sample was evenly distributed. Based on these findings, we offer survey design and data collection recommendations for paid online panel contexts.

The rest of the paper is organized as follows: Section 2 provides a theoretical overview of survey framing dynamics and the hypotheses. Section 3 details the research design, development of the two surveys, data collection procedures, and sample details. Section 4 reports the results on the differences in data quality between the survey samples. Section 5 discusses the study's methodological contributions and limitations, provides recommendations for researchers, and offers suggestions for future research.

## Theoretical background and hypotheses

Data quality is not something researchers should consider only after data collection. Instead, quality assurance mechanisms must be directly embedded within the survey design before data collection begins (Arndt et al., 2022). Researchers also cannot delegate data quality responsibilities to panel companies. These companies compete for clients who often have limited budgets and tend to prioritize obtaining as many responses as possible, as quickly as possible, to meet minimum sample size requirements. Although panel providers manage their respondent pools (e.g., by removing panelists who consistently fail attention trap questions, ATQs), this competitive environment does not foster optimal data quality.

As a result, researchers must invest significant resources into survey design to ensure the integrity and quality of the collected data. Methodology literature suggests several remedies and methods for achieving high data quality. These include keeping the survey and items concise to avoid fatigue (Arndt et al., 2022), including ATQs to identify inattentive respondents (Meade & Craig, 2012), and emphasizing that responses are anonymous and used solely for research purposes to encourage honest rather than socially desirable responses (Podsakoff et al., 2024). However, even the most carefully designed surveys can yield poor quality or irrelevant data if they are completed by respondents who lack the necessary background or motivation. This challenge is particularly critical in IS research requiring respondents with specific IS domain knowledge or users of particular IT environments, who are hard to reach via online panels.

Therefore, a more fundamental issue lies in determining who is allowed to take the survey in the first place. Data cleaning typically catches only blatant issues, potentially leaving attentive but unqualified respondents in the dataset. *This concern is amplified when internal panel screeners lack the granularity needed to restrict access to the survey to the intended target sample, particularly when the goal is to gather insights from individuals with genuine domain experience rather than general attitudes* (Belliveau & Yakovenko, 2022; Pickering & Blaszczynski, 2021). To address this, researchers have proposed using hidden eligibility criteria to screen in only relevant respondents (Sharpe Wessling et al., 2017).

This leads to a key survey design decision: whether to use explicit or hidden framing when recruiting participants (Chandler & Paolacci, 2017; Sharpe Wessling et al., 2017). Explicit survey framing involves openly stating the topic and target population of the survey during the recruitment phase or at the beginning of the survey. For example, our survey recruitment used the title "Attitudes Toward Electrification of Non-Road Work Vehicles: A Survey for Current Operators", which transparently disclosed the survey's eligibility criteria and enabled participants to opt in based on their self-perceived eligibility and interest. This approach reduces ambiguity and cognitive load for respondents, helping them align their expectations with survey content. It also fosters trust between the survey panelist and the survey administrator and is generally more convenient and less expensive (Sharpe Wessling et al., 2017). Panel platforms often promote their participant pools as honest and capable of delivering high-quality data. Comparative research also suggests that certain platforms have more honest respondents (Peer et al., 2022), which may lead researchers to forego more rigorous screening and opt for transparent, explicit framing.

However, using transparent eligibility criteria may also encourage savvy survey panel participants to misrepresent their qualifications in order to participate and receive a reward (Chandler & Paolacci, 2017; Hyman & Sierra, 2012). Financial rewards are a key motivator for many online survey panel members (Pickering & Blaszczynski, 2021), although altruism, scientific curiosity, and professional interest also play a role (Keusch et al., 2014). Prior studies have shown that eligibility misrepresentation and impostor participation are widespread in paid online survey panels when the survey topic is explicitly disclosed (Chandler & Paolacci, 2017; Sharpe Wessling et al., 2017), suggesting that respondents without actual experience related to the survey topic may be included under explicit framing. These impostor participants

can systematically bias results and distort the prevalence or dynamics of the phenomena under investigation (Chandler & Paolacci, 2017; Pickering & Blaszczynski, 2021).

In contrast, hidden framing disguises the survey's true focus using a broader initial description. For example, our second survey used the title "Attitudes Toward Electrification of Work Vehicles: A Survey for Current Operators", which left the topic open to all work vehicle users. Participants are then typically presented with a mix of actual and decoy screening questions, such as whether they operate vehicles on primary roads or in off-road environments (actual screener), and whether they perform maintenance themselves (decoy question). This obfuscation makes it harder for participants to infer which question-and-answer combinations grant access to the full survey and its associated reward (Chandler & Paolacci, 2017; Sharpe Wessling et al., 2017). While this approach may yield more relevant respondents, it is more complex and resource-intensive, as screened-out participants must also be compensated and may raise complaints about being excluded (Sharpe Wessling et al., 2017), resulting in added work for the survey administrator in resolving issues and responding to inquiries.

Given these considerations and tradeoffs, a key expected difference between explicit and hidden framing lies in the proportion of participants who are screened out. When both samples are drawn from the same pool using identical internal panel screeners, eligibility must be communicated and verified through one of two mechanisms. In the explicit framing condition, the survey topic is clearly stated, which draws interest from genuinely eligible participants and impostors, typically resulting in fewer exclusions due to self-selection. In contrast, hidden framing employs a broader initial description, allowing screener questions and the survey's focus to remain obfuscated. This approach is likely to yield a higher screen-out rate, particularly because non-road work vehicle operators represent a low-incidence population within the broader panel. However, it may also produce a more relevant final sample. We therefore hypothesize:

H1: The explicit framing sample will have fewer screened out participants compared to the hidden framing sample.

The explicit framing approach, combined with transparent screeners, may leave researchers more open to being "gamed" by professional survey-takers (Arndt et al., 2022). Many paid respondents are known to be inattentive and inconsistent, often speeding through surveys to maximize their income generation potential (Berinsky et al., 2014). Inattentiveness can critically impact data quality, which is why it is commonly measured using multiple ATQs (Berinsky et al., 2014; Meade & Craig, 2012).

However, not all inattentive behavior is equal. Participants on platforms like Prolific are typically more experienced than casual survey-takers and are familiar with common survey design mechanisms, including ATQs (Peer et al., 2022). Because failing ATQs may result in denied rewards or even eventual removal from the platform after multiple failures, these respondents are motivated to remain strategically attentive to such checks, even if they are not part of the target demographic (Sharpe Wessling et al., 2017).

Consequently, while some impostors may rush through the survey and provide low-effort responses, their familiarity with instructional checks and the repercussions of failing them means they are still likely to pass ATQs. Meanwhile, respondents screened in via hidden criteria are more likely to have a genuine interest in the topic, which may enhance attentiveness. Given these counterbalancing influences, we hypothesize:

H2: There will be no difference in attention trap passing rates between the framing samples.

Although impostors in the explicit framing sample are likely to remain vigilant toward ATQs, they are also likely to spend less time and cognitive effort on substantive survey questions, as responding accurately would require them to fabricate expertise (Sharpe Wessling et al., 2017). While a subset of impostors may attempt to fake expertise convincingly, the fixed reward structure incentivizes most to engage in satisficing, that is, providing quick, "good enough" answers (Barge & Gehlbach, 2012; Meade & Craig, 2012). Because the survey theme holds less intrinsic relevance to impostors, they are more likely to prioritize speed over deep engagement to maximize income (Hyman & Sierra, 2012), resulting in a faster average response time. In contrast, because the hidden framing sample likely consists only of respondents with actual domain experience, they are likely to spend more time reflecting on the questions. Thus, we hypothesize that:

H3: The average time spent on completing the survey will be shorter for the explicit framing sample compared to the hidden framing sample.

When survey rewards depend on successfully completing the survey, impostors face an evaluation threat from the survey administrator, where the reward for the survey or participation in future surveys may be denied if they appear untrustworthy or inattentive. In such scenarios, respondents may inflate socially desirable participant engagement traits, which we operationalize as interest, diligence, and integrity (Meade & Craig, 2012; Schlenker, 2008). To avoid extra scrutiny from survey administrators, the explicit framing sample, which likely includes more impostors, is thus not expected to differ significantly on these measures. This is especially plausible given that respondents had already encountered ATQs, which can heighten socially desirable responding (Clifford & Jerit, 2015). Therefore, we hypothesize the following:

H4: There will be no difference between self-reported engagement measures between the samples.

Respondents primarily motivated by financial rewards, rather than genuine interest in the survey topic, are more likely to engage in low-effort response styles (Barge & Gehlbach, 2012; Chandler & Paolacci, 2017). This can compromise data quality by reducing thoughtful engagement and increasing socially desirable or normatively expected answers. Traditionally, data quality has been assessed using metrics such as construct reliability, convergent and discriminant validity, model fit, and CMV (Chen, 2007; Fornell & Larcker, 1981; Hair et al., 2014; Henseler et al., 2015; Williams et al., 2010). Although various screening and framing methods have been proposed (Arndt et al., 2022), the differences in data quality between self-selected samples and those screened in using concealed criteria remain unclear.

Hidden framing combined with concealed screening is likely to yield a sample that better matches the target population, resulting in more relevant responses grounded in actual domain experience. These respondents are more likely to engage critically with the survey content, as they are more familiar with the actual work condition challenges. In contrast, imposters in the explicit framing sample, who may lack relevant experience, are less likely to invest cognitive effort and more likely to provide socially desirable answers aligning with perceived expectations (Barge & Gehlbach, 2012; Kwak et al., 2019; Protzko et al., 2019).

Differences in respondent relevance and engagement may also influence CMV. If CMV is present, the explicit framing sample is likely to have unevenly distributed CMV due to a higher prevalence of socially desirable response styles (e.g., acquiescing to positive attitude or sustainability-related questions) (Podsakoff et al., 2024; Protzko et al., 2019). We expect the hidden framing sample to exhibit higher construct reliability, as well as stronger convergent and discriminant validity, resulting in non-invariant measurement models across samples. Specifically, the hidden framing sample is also anticipated to show a more uniform or lower pattern of CMV across items. Thus, we hypothesize:

H5: The hidden framing sample will exhibit better data quality metrics compared to the explicit framing sample.

## **Methodology and survey design**

Using two survey designs (explicit versus hidden framing), we examine how a survey's topic transparency influences respondent screening and downstream data quality (Sharpe Wessling et al., 2017). We chose non-road work vehicles as the study context because Prolific's internal screening mechanisms could not precisely target our population (e.g., no screeners were available to identify whether panelists operated vehicles at work or in which contexts). This context typifies IS studies that need data from actual users of specific technologies (e.g., employees using machine-learning applications at work), where internal platform-level screeners often lack the granularity needed for precise selection. Moreover, the domain targets a low-incidence, specialized professional population, unlike users of mass-market, ubiquitous IS, in which ineligible respondents could misinform safety-critical, capital-intensive organizational decisions.

In the explicit framing, the title of the survey recruitment message ("Attitudes Toward Electrification of Non-Road Work Vehicles: A Survey for Current Operators) explicitly stated who the survey was intended for, with a brief description about the study context. Participants were then able to open the introduction page of the survey where the topic was explained in more detail. At the end of the introduction page, respondents were asked to confirm whether they operated non-road work vehicles as described in the page. In contrast, the hidden framing had a broader description for the topic of the survey, encompassing all types of work vehicle operators ("Attitudes Toward Electrification of Work Vehicles: A Survey for Current Operators"). The first page these respondents entered included the previously mentioned hidden screener question set, including the actual screener questions and decoy questions intended to obfuscate the role of

the questions. The screener questions were: “Do you operate vehicles at work?” (answering “Yes” led to being screened in, in combination with the latter question) and “Do you primarily operate your vehicle on main roads or off-road environments?” (answering “Primarily in non-road or off-road environments (e.g., warehouses, construction sites, industrial facilities, natural terrain, ports or airports etc.)” led to being screened in). The decoy questions were: “Do you have a driver’s license?”, “Do you travel long distances with work vehicles for your job?”, and “Do you perform maintenance on work vehicles yourself?”

If the respondents answered correctly to the screeners, they were admitted to the full survey and were eligible for the full reward upon completion. A reward of £2.75 ( $\approx$ €3.31) was paid for completing the survey. This was rated as a “Great!” reward rate by Prolific, indicating that the survey was likely to garner a lot of interest from the panelists, as well as more fraudulent self-selection (Chandler & Paolacci, 2017). Those who provided wrong answers to the more complex screener-set were screened out and paid a small reward for their time (£0.15–£0.5, depending on Prolific’s dynamic recommendation based on response time). The respondents were allowed to answer the survey using either a desktop/laptop or a tablet, as smartphone respondents have been found to provide lower-quality data (Peer et al., 2022) and are more likely to drop out if the survey is not optimized for smartphones. The respondents were informed that the survey would take approximately 15 minutes to complete. They were required to answer all questions, which were measured using a 7-point Likert scale ranging from “Strongly disagree” to “Strongly agree”.

After the introduction page, the rest of the surveys were identical for both framing conditions. First, background and demographic information were collected. The following page included independent variables examining EWV adoption benefits and challenges, as well as the first ATQ. The next page included additional independent variables examining personal impact and organizational support perceptions. The following page included the dependent variables examining attitudes toward EWV adoption and the second ATQ. The final page included items measuring CMV and participant engagement, specifically diligence, interest, and integrity (Meade & Craig, 2012; Schlenker, 2008).

Ten constructs with different focuses were selected for our analysis. We included positively valenced constructs that are more prone to socially desirable response styles (Podsakoff et al., 2012), such as sustainability, as well as constructs requiring domain knowledge, such as range sufficiency. The intention to use scale (Venkatesh et al., 2003) was adapted to focus on employee perceptions toward organizational adoption intentions. Colleague opinion (Kim & Kankanhalli, 2009) and attitude (Venkatesh et al., 2003) scales were rephrased for the EWV context. The resistance scale (Kim & Kankanhalli, 2009) was adapted to assess attitudinal resistance. The inertia scale (Polites & Karahanna, 2012) was adapted to assess employee preferences toward traditional work vehicles. The techno-overload scale (Tarafdar et al., 2007) was rephrased to assess employee perceptions of EWV use complexity resulting in increased workload. The sustainability scale (Fang & Li, 2022) was adapted to assess employees’ beliefs about the sustainability of EWVs compared to traditional vehicles. The range sufficiency scale (Schmalfuß et al., 2017) was adapted to assess how comfortable employees were with the range of EWVs, while the low noise preference scale (Schmalfuß et al., 2017) gauged whether the respondents would enjoy the quietness of EWVs. The CMV marker variable (MV) assessed attitude toward the color blue (ATCB), which has been suggested to be an ideal MV, as it is theoretically unrelated to the substantive variables (Miller & Simmering, 2023).

To proactively mitigate CMV, the respondents were informed that their responses would be anonymous and used only for research purposes (Podsakoff et al., 2024). The latent MV was used in post-hoc CMV analysis (Miller & Simmering, 2023; Williams et al., 2010). To examine what impact the different survey framings had on data quality, the descriptive demographic and construct statistics were first calculated using IBM SPSS version 29. This was followed by confirmatory factor analysis (CFA) in IBM SPSS Amos version 28. We assessed construct reliability using composite reliability (CR), convergent validity using average variance extracted (AVE), and discriminant validity using both the Fornell–Larcker criterion (comparing  $\sqrt{AVE}$  to inter-construct correlations) and the heterotrait-monotrait (HTMT) ratio of correlations (Fornell & Larcker, 1981; Hair et al., 2014; Henseler et al., 2015). Additionally, we examined model fit and measurement invariance between the samples (Chen, 2007; Hu & Bentler, 1999).

## **Data collection**

The data were collected from Prolific between November 2024 and December 2024. We used non-probabilistic, or convenience sampling (Hays et al., 2015), due to the study context, as operators of non-road work vehicles are unlikely to be evenly distributed across demographic groups. The data were collected

in several stages, beginning with 250 spots in the pilot version of each survey framing. Following this, the survey was launched in multiple batches to collect the full sample based on our estimated remaining budget. Considering that screened-out participants also had to be compensated for their time, achieving identical sample sizes across conditions was expected to be challenging. However, we were able to obtain sufficiently similar sample sizes for joint analysis.

Participants were first filtered using Prolific's internal screeners. The specific screeners included the participant owning a driver's license or a learner's permit, being employed full-time or part-time, originating from Europe, USA, Canada, New Zealand, or Australia, and being employed in an industry where non-road vehicles are likely to be used (e.g., manufacturing, warehousing, or construction). To ensure a pool of naive respondents unfamiliar with the survey (Chandler & Paolacci, 2017), participants who had already completed one of the framing surveys were excluded from participating in the other.

Overall, 1,753 complete survey responses were recorded: 1,045 for the explicit framing survey and 708 for the hidden framing survey. Despite our best efforts to exclude participants who had completed another version of the survey, 11 respondents with identical ProlificIDs completed both surveys. After removing these duplicates and one respondent with missing values, the final sample is 1,730. No additional data cleaning procedures (e.g., removing responses due to failing ATQs or straight lining) were performed. The background characteristics of the samples can be seen in Table 1.

	Explicit framing (n = 1,035)		Hidden framing (n = 695)	
<b>Gender</b>	<b>Frequency</b>	<b>%</b>	<b>Frequency</b>	<b>%</b>
Male	766	74	501	72.1
Female	261	25.2	191	27.5
Other	8	0.8	3	0.4
<b>Age</b>				
From 18 to 34 years	522	50.4	399	57.4
From 35 to 54 years	423	40.9	246	35.4
From 55 to 74 years	90	8.7	50	7.2
<b>Education</b>				
Less than high school	8	0.8	6	0.9
Graduated high school	161	15.6	109	15.7
Trade/technical school	94	9.1	52	7.5
Some college, no degree	149	14.4	121	17.4
Associate degree	54	5.2	51	7.3
Bachelor's degree	420	40.6	244	35.1
Advanced degree	149	14.4	112	16.1
<b>Country</b>				
United Kingdom	352	34.0	179	25.8
United States	346	33.4	321	46.2
Canada	78	7.5	32	4.6
Other	259	25.1	163	23.4

**Table 1. Respondents' Demographic Background**

## Survey framing comparison results

The means (M), standard deviations (SD), and item wordings are reported in Table 2. The standardized CFA loadings ( $\lambda$ ) were predominantly above the suggested .707 cutoff, indicating that the constructs explained at least 50% of the variance in their indicators (Hair et al., 2014), although a few items (RESI2

and ATTI1) had loadings that were below this threshold. Given the high sample sizes, these items were retained for the subsequent analyses, as they still measure unique variance and contribute to the content validity of the constructs. Additionally, a minimum of three items is recommended for model identification (Hair et al., 2014). All standardized factor loadings were statistically significant ( $p < .001$ ).

<b>Construct</b>	<b>Item wording</b>	<b>M (SD)</b>	<b><math>\lambda</math></b>
Organizational adoption intention (Venkatesh et al., 2003)	OFUS1: I predict that our organization will use electric work vehicles in the future	5.42 (1.36) 5.28 (1.50)	.797 .871
	OFUS2: Our organization plans to use electric work vehicles in the future	5.09 (1.42) 5.03 (1.57)	.895 .926
	OFUS3: Our organization intends to use electric work vehicles in the future	5.09 (1.44) 5.04 (1.56)	.895 .931
Colleague opinion (Kim & Kankanhalli, 2009)	COOP1: Most of my colleagues think switching to electric work vehicles is a good idea	4.68 (1.45) 4.58 (1.57)	.851 .892
	COOP2: My peers support switching to electric work vehicles	4.70 (1.36) 4.62 (1.49)	.812 .882
	COOP3: Most people I work with encourage switching to electric work vehicles	4.53 (1.44) 4.37 (1.57)	.841 .864
Resistance (Kim & Kankanhalli, 2009)	RESI1: I am reluctant to switch to using electric work vehicles	3.0 (1.70) 3.11 (1.78)	.777 .830
	RESI2: I am unwilling to switch to using electric work vehicles	2.63 (1.71) 2.82 (1.83)	.674 .664
	RESI3: I oppose switching to using electric work vehicles	2.57 (1.59) 2.79 (1.80)	.857 .842
	RESI4: I do not agree with switching to using electric work vehicles	2.58 (1.56) 2.74 (1.71)	.847 .857
Attitude (Venkatesh et al., 2003)	ATTI1: Electric work vehicles would make work more interesting	4.71 (1.40) 4.71 (1.48)	.686 .739
	ATTI2: Working with electric work vehicles would be fun	5.16 (1.25) 5.11 (1.35)	.812 .838
	ATTI3: I would enjoy working with electric work vehicles	5.33 (1.21) 5.30 (1.33)	.835 .872
Inertia (Polites & Karahanna, 2012)	INER1: I prefer traditional work vehicles because they are stress-free	3.52 (1.62) 3.48 (1.68)	.817 .795
	INER2: I prefer traditional work vehicles because using them is more comfortable for me	3.66 (1.61) 3.80 (1.70)	.877 .849
	INER3: I prefer traditional work vehicles because I have been working with them for so long	3.71 (1.66) 3.79 (1.76)	.870 .840
	INER4: I prefer traditional work vehicles because I have used them regularly in the past	3.83 (1.67) 3.91 (1.71)	.853 .854
Techno-overload (Taraifdar et al., 2007)	OVER1: The complexity of using electric vehicles would force me to do more work than I can handle	3.03 (1.59) 2.98 (1.60)	.817 .803
	OVER2: The complexity of using electric vehicles would leave me less time to focus on my actual work	3.03 (1.62) 3.03 (1.66)	.823 .838

	OVER3: I would have a higher workload because of the complexity of electric vehicles	3.1 (1.61) 3.13 (1.66)	.792 .791
Sustainability (Fang & Li, 2022)	SUST1: Using electric work vehicles helps conserve more resources than traditional vehicles	5.17 (1.33) 5.15 (1.41)	.794 .821
	SUST2: Using electric work vehicles is more sustainable than traditional vehicles	5.18 (1.42) 5.10 (1.52)	.759 .822
	SUST3: Using electric work vehicles is more efficient in terms of resource utilization than traditional vehicles	5.04 (1.30) 5.07 (1.41)	.733 .819
	SUST4: Using electric work vehicles is more environmentally friendly than using traditional vehicles	5.50 (1.37) 5.38 (1.40)	.754 .813
Range sufficiency (Schmalfuß et al., 2017)	RASU1: I am comfortable with the range of electric work vehicles	5.02 (1.37) 4.89 (1.50)	.818 .819
	RASU2: The range of electric work vehicles is satisfactory	4.84 (1.38) 4.74 (1.47)	.798 .831
	RASU3: The range of electric work vehicles is sufficient for my work mobility needs	5.04 (1.33) 4.92 (1.50)	.766 .797
Low noise preference (Schmalfuß et al., 2017)	LONO1: The low noise level of electric work vehicles would make driving more enjoyable	5.40 (1.35) 5.27 (1.45)	.758 .781
	LONO2: I would like the quietness of electric work vehicles	5.51 (1.32) 5.42 (1.47)	.785 .836
	LONO3: I would perceive the low noise level of electric work vehicles as pleasant	5.56 (1.24) 5.50 (1.31)	.803 .772
Attitude toward the color blue (Miller & Simmering, 2023)	BLUE1: Blue is a beautiful color	5.72 (1.20) 5.78 (1.10)	.968 .934
	BLUE2: Blue is a nice color	5.78 (1.15) 5.84 (1.06)	.963 .932
	BLUE3: I like the color blue	5.76 (1.18) 5.81 (1.11)	.940 .918
	BLUE4: I think blue is a pretty color	5.70 (1.22) 5.80 (1.08)	.936 .935

**Table 2. Item Wordings, Means, Standard Deviations, and Standardized Loadings  
(cell format: explicit framing value above, hidden framing value below)**

### Screen-out rates, attentiveness, and self-reported participant engagement

For the explicit framing, a total of 1,711 interactions with the survey were recorded. Out of these, 1,045 (61.08%) responses were approved after completing the survey, 463 (27.06%) returned their spot reservation after opening the survey, and 203 (11.87%) screened themselves out by answering “No” to whether they operated non-road vehicles in the introduction page. For the hidden framing, a total of 5,511 interactions with the survey were recorded. From these, 708 (12.85%) completed responses were approved, 486 (8.82%) returned their spot, and 4,317 (78.33%) were screened out using the in-survey screening questions (1: “Do you operate vehicles in your work?” and “Where do you mainly operate your vehicle?”). Respondents had to answer “Yes” to the first screener question and “Primarily in non-road or off-road environments (e.g., warehouses, construction sites, industrial facilities, natural terrain, ports or airports etc.)” to the second screener question; otherwise, they were screened out of the survey. A considerable difference between the screen-out rates can be observed, and a chi-square test ( $\chi^2(2) = 2,409.90$ ,  $p < .001$ ) showed a significant relationship between the framing condition and the screen-out rate. This result provides strong support for H1.

The survey included two ATQs to measure attentiveness: (1) "Please select Strongly disagree as your answer for this statement to show that you are paying attention" on page 3, and (2) "To confirm that you are still paying attention, please select Strongly disagree as your response to this statement" on page 5. In the explicit framing sample, 95.4% passed the first ATQ, 96% the second, and 93% passed both. In the hidden framing sample, 93.8% passed the first ATQ, 96.5% passed the second, and 92.8% passed both. Chi-square tests did not show a statistically significant relationship between the framing condition and ATQ success rates for the first ( $\chi^2(1) = 2.003$ ,  $p = .157$ ), second ( $\chi^2(1) = 0.297$ ,  $p = .586$ ), or both ( $\chi^2(1) = 0.036$ ,  $p = .850$ ) ATQs, indicating no statistically significant differences in attention-trap pass rates between the samples. This result supports H2. To assess the raw data quality obtained using the different framing conditions, those who failed the ATQs were retained in the subsequent analyses.

To assess differences in survey completion times, we first identified outliers by calculating the Z scores for the completion times to identify respondents who may have, for instance, left the survey open for several hours before finally completing it. We identified two respondents with a Z score  $> 2$  that were excluded from the following t-test (Hair et al., 2014). For the explicit sample, the mean response time was 14.18 minutes ( $SD = 7.61$  minutes), with a median of 12.45 minutes, a minimum 2.72 minutes, and a maximum of 68.97 minutes. For the hidden sample, the mean response time was 15.36 minutes ( $SD = 6.50$  minutes), with a median of 13.18 minutes, a minimum 3.05 minutes, and a maximum of 55.02 minutes. An independent samples t-test comparing the sample response times was statistically significant ( $p = .003$ ,  $p < .01$ ), confirming there was a significant difference in means between the samples, supporting H3.

We measured participant engagement by assessing self-reported interest in the topic and self-assessed diligence (i.e., whether participants answered the survey attentively) (Meade & Craig, 2012), as well as integrity (i.e., whether one believes their principles should not be sacrificed for financial gain) using four-item scales (Schlenker, 2008). Construct scores were calculated by averaging the items. An independent samples t-test comparing the means between the samples with a 90% confidence interval for two one-sided tests (TOST) was conducted. We set the smallest effect size of interest (SESOI) at  $\Delta_L = -0.3$  and  $\Delta_U = 0.3$  Likert scale points (Lakens et al., 2018). The analysis indicated statistical equivalence for interest ( $M_E = 5.69$ ,  $SD_E = 1.04$  versus  $M_H = 5.74$ ,  $SD_H = 0.98$ ,  $\Delta_L = -0.134$ ,  $\Delta_U = 0.030$ ), diligence ( $M_E = 6.41$ ,  $SD_E = 0.7$  versus  $M_H = 6.38$ ,  $SD_H = 0.69$ ,  $\Delta_L = -0.026$ ,  $\Delta_U = 0.086$ ), and integrity ( $M_E = 5.09$ ,  $SD_E = 1.12$  versus  $M_H = 5.13$ ,  $SD_H = 1.11$ ,  $\Delta_L = -0.130$ ,  $\Delta_U = 0.050$ ), as the  $\Delta_L$  and  $\Delta_U$  intervals fell within the SESOI, supporting H4.

### **Construct reliability, convergent validity, and discriminant validity**

Across both measurement models (explicit and hidden framing samples), we evaluated construct reliability ( $CR \geq .7$ ), convergent validity ( $AVE \geq .5$ ), and discriminant validity ( $\sqrt{AVE} >$  any inter-construct correlation;  $HTMT < .85$ ) (Fornell & Larcker, 1981; Hair et al., 2014; Henseler et al., 2015). The results obtained using the Master Validity plugin (Gaskin et al., 2019) indicate that both samples meet the criteria for construct reliability, convergent validity, and discriminant validity (Table 3). This result does not support H5.

COOP	LONO	RASU	ATII
.093 .067	.110 .108	.115 .098	.148 .091
.332 .322	.343 .377	.365 .358	.364 .415
.499 .581	.506 .546	.487 .588	.576 .630
.598 .635	.308 .388	.461 .521	.470 .553
.058 .027	.264 .249	.196 .202	.164 .129
.325 .334	.394 .398	.373 .400	.417 .456
<b>.835</b> <b>.879</b>	.327 .388	.503 .570	.554 .642
.383 .449	<b>.782</b> <b>.797</b>	.357 .413	.519 .558
.584 .648	.430 .489	<b>.794</b> <b>.815</b>	.491 .566
.646 .721	.633 .660	.602 .669	<b>.780</b> <b>.818</b>

Construct	CR	AVE	BLUE	INER	SUST	OFUS	OVER	RESI
BLUE	.975 .962	.907 .864	<b>.952</b> <b>.930</b>	.028 .000	.128 .094	.062 .071	.023 .030	.019 .018
INER	.916 .902	.731 .697	.031 .000	<b>.855</b> <b>.835</b>	.366 .447	.353 .371	.622 .577	.660 .634
SUST	.845 .890	.578 .670	.141 .102	-.413 -.499	<b>.760</b> <b>.819</b>	.430 .504	.173 .262	.426 .472
OFUS	.898 .935	.746 .828	.060 .076	-.379 -.399	.473 .544	<b>.863</b> <b>.910</b>	.212 .181	.400 .432
OVER	.852 .852	.658 .658	.027 -.029	.697 .649	-.198 -.302	-.221 -.197	<b>.811</b> <b>.811</b>	.641 .609
RESI	.870 .877	.627 .643	-.013 -.019	.736 .721	-.501 -.559	-.435 -.474	.736 .701	<b>.792</b> <b>.802</b>
COOP	.873 .911	.697 .773	.097 .073	-.372 -.361	.575 .648	.669 .688	-.063 -.035	-.387 -.391
LONO	.825 .839	.612 .635	.124 .120	-.389 -.433	.602 .632	.345 .434	-.314 -.295	-.471 -.484
RASU	.837 .856	.631 .665	.126 .109	-.417 -.406	.574 .673	.523 .582	-.225 -.234	-.439 -.472
ATTI	.823 .858	.609 .670	.154 .104	-.453 -.502	.680 .722	.539 .619	-.243 -.199	-.539 -.582

**Table 3. Construct Reliability, Convergent Validity, and Discriminant Validity (bold diagonal =  $\sqrt{AVE}$ ; lower triangle = inter-construct correlations for the Fornell–Larcker criterion; upper triangle = HTMT; cell format = explicit framing value above, hidden framing value below)**

## Model fit

The chi-square test for both measurement models was significant ( $p < .05$ ), indicating poor fit. However, this is typical with large sample sizes (Schermelleh-Engel et al., 2003), which is why model fit was also assessed using the comparative fit index (CFI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and the ratio between the chi-square ( $\chi^2$ ) and degrees of freedom (df) (Hair et al., 2014). Cutoff values for excellent model fit are CFI  $> .95$ , SRMR  $< .08$ , RMSEA  $< .06$ , and  $1 < \chi^2/df < 3$  (Hu & Bentler, 1999). Both samples indicated excellent model fit (Table 4). This does not support H<sub>5</sub>.

Model fit measure	$\chi^2$	df	$\chi^2/df$	CFI	SRMR	RMSEA
Explicit framing sample	1,234.831	482	2.562	.971	.040	.039
Hidden framing sample	911.017	482	1.890	.977	.038	.036

**Table 4. Measurement Model Fit for Both Samples**

### **Configural, metric, and scalar invariance**

We tested for configural invariance by estimating a model with two groups (explicit,  $n = 1,035$  and hidden,  $n = 695$ ). The configural model showed excellent model fit ( $CFI = .973$ ,  $SRMR = .0391$ ,  $RMSEA = .027$ ) (Hu & Bentler, 1999). Configural invariance was thus established, indicating equivalent factor structures and a similar conceptual understanding of the items across both samples (Chen, 2007; Milfont & Fischer, 2010). Metric invariance was assessed by fixing all factor loadings equal across both samples. A chi-square difference test indicated that the models were not invariant ( $p < .01$ ). However, this is typical with larger sample sizes, which is why we compared the differences in model fit indices as suggested by Chen (2007) using the cutoffs of  $|\Delta CFI| \leq .01$ ,  $|\Delta RMSEA| \leq .015$ , and  $|\Delta SRMR| \leq .03$ . According to these criteria, the models were invariant, establishing metric invariance. Although factor loadings varied between the samples (Table 2), these differences were not statistically significant. To assess scalar invariance, we constrained the item intercepts to be equal across both samples. The chi-square difference test once again indicated that the models were not invariant ( $p < .05$ ). However, the differences in model fit indices were negligible (Table 5). Scalar invariance was thus also established. This does not support H5.

Model	Configural	Metric	Scalar
<b>X<sup>2</sup></b>	2,145.871	2,205.830	2,255.476
<b>df</b>	964	998	1032
<b>CFI</b>	.973	.973	.972
<b>SRMR</b>	.0391	.0462	.0464
<b>RMSEA</b>	.027	.026	.026
<b>ΔX<sup>2</sup></b>	-	59.959	49.646
<b>Δdf</b>	-	34	34
<b>sig.</b>	-	$p = .004$	$p = .041$
<b> ΔCFI </b>	-	0	.001
<b> ΔSRMR </b>	-	.0071	.0002
<b> ΔRMSEA </b>	-	.001	0

**Table 5. Measurement Invariance Testing Results**

### **Common method variance and bias**

We utilized the CFA marker technique for CMV analysis (Williams et al., 2010). We used the attitude toward the color blue as a theoretically unrelated latent MV (Miller & Simmering, 2023). The results of the analysis for both framing conditions are reported in Table 6. We first specified a Baseline model in which the latent MV was allowed to correlate with all substantive constructs, estimating its unstandardized factor loadings and error variances; the factor loadings and error variances of the latent MV were then fixed to these obtained estimates in all subsequent models. Following this, we specified the Method-C model in which we constrained the correlations between the latent MV and all substantive constructs to zero and added factor loadings between the latent MV and every substantive item, fixing these loadings equal across items. In the Method-U model, the factor loadings between the latent MV and substantive items were allowed to be freely estimated. Finally, to assess whether any potential CMV resulted in common method bias (CMB), we compared either the Method-C or Method-U model, depending on which had a better fit with the data, to the Method-R model, in which the correlations between the substantive constructs were fixed to those obtained from the Baseline model.

For the explicit framing sample, the chi-square difference test comparing the Baseline and Method-C models was statistically significant ( $p < .001$ ), suggesting that CMV was present. The comparison between the Method-C and Method-U models was also statistically significant ( $p < .05$ ), indicating that CMV was congeneric, that is, unevenly distributed across the substantive items (Williams et al., 2010). Finally, the comparison between the Method-U and Method-R models was statistically insignificant ( $p > .05$ ), suggesting that the unevenly distributed CMV does not affect the correlations between the constructs.

For the hidden framing sample, the chi-square difference test comparing the Baseline and Method-C models was statistically significant ( $p = .001$ ), suggesting that CMV was present in this sample as well. However, the comparison between the Method-C and Method-U models showed a statistically insignificant result ( $p > .05$ ), indicating non-congeneric CMV, where the CMV is uniformly distributed across the substantive items (Williams et al., 2010). We thus retained the Method-C model and compared it against the Method-R model, which showed a statistically insignificant result ( $p > .05$ ), indicating that the evenly distributed CMV does not result in CMB by affecting the correlations between the substantive constructs.

In summary, the samples had differing CMV profiles. The explicit sample exhibited unevenly distributed CMV, impacting positively valenced items more than others, whereas the hidden sample exhibited evenly distributed CMV. In contrast to earlier H5-related analysis, the CMV results provide partial support for H5.

<b>Model (explicit)</b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>CFI</b>	<b><math>\Delta\chi^2</math></b>	<b><math>\Delta df</math></b>	<b>p</b>
Baseline	1,276.494	499	.970	-	-	-
Method-C	1,242.305	498	.971	34.189	1	< .001
Method-U	1,194.349	469	.972	47.956	29	.015
Method-R	1,195.002	505	.973	0.653	36 (U & R)	> .05
<b>Model (hidden)</b>	<b><math>\chi^2</math></b>	<b>df</b>	<b>CFI</b>	<b><math>\Delta\chi^2</math></b>	<b><math>\Delta df</math></b>	<b>p</b>
Baseline	927.859	499	.977	-	-	-
Method-C	917.109	498	.977	10.75	1	.001
Method-U	901.367	469	.976	15.742	29	> .05
Method-R	917.264	534	.979	0.155	36 (C & R)	> .05

**Table 6. CFA Marker Analysis Results**

## Discussion

Our study examined two survey framing approaches in a paid online survey panel context: (1) explicit topic disclosure with participant self-screening, and (2) hidden survey framing with screening criteria concealed from participants, to assess their effects on screening rates and data quality. Using these approaches, the hidden framing sample is expected to include only the target population, whereas the explicit framing sample is likely to contain a mix of genuine respondents with domain experience and impostors (Chandler & Paolacci, 2017; Sharpe Wessling et al., 2017). Our findings are particularly relevant for researchers seeking to collect data from individuals with genuine experience related to the study topic, rather than general attitudes toward broader issues. While especially applicable to paid online survey panels, these insights may also extend to other contexts, such as crowdsourced surveys that offer no financial incentives.

### Methodological contributions

The results have several implications and tradeoffs. First, screen-out rates differed sharply between the framing conditions (11.87% for explicit, 78.33% for hidden), suggesting the presence of potential impostors in the explicit framing sample. Although this gap is partly a design effect (i.e., due to use of concealed eligibility criteria in the hidden condition), its magnitude remains concerning. The finding provides further support for the use of concealed pre-screening methods, such as conducting a separate screener study or incorporating a set of screener questions prior to the main survey, rather than allowing participants to self-select into the survey (Chandler & Paolacci, 2017; Sharpe Wessling et al., 2017). Such an approach is more likely to yield a sample consisting of respondents with domain experience.

Second, explicit framing is less expensive because screened-out participants do not need to be compensated. Depending on the specificity of the topic, hidden framing can result in a high number of exclusions, making data collection costly both monetarily and timewise. The explicit framing approach is more convenient for both researchers and respondents, as participants can self-select into surveys that match their interest, but risks more irrelevant respondents (Sharpe Wessling et al., 2017). In contrast, using screeners to select participants based on concealed criteria may yield more relevant responses but requires additional budget for the screened-out participants and extra effort from the survey administrator to handle inquiries from

panel members about their removal from the survey. However, using screener questions to ensure respondent relevancy is not a silver bullet for achieving high data quality. Once participants have been screened into the survey, they may still proceed to answer as rapidly as possible to collect the reward, even if the theme is relevant to them.

Third, we found no differences between the samples for most data quality metrics, specifically in rates of passing attention traps, construct reliability, or convergent and discriminant validity. We also established measurement invariance up to scalar invariance when examining both samples. However, we identified a difference in the CMV profile of the samples. The explicit framing sample likely included more impostors with no actual domain experience. This likely resulted in CMV being unevenly distributed across the items for the explicit framing sample, focusing on positively valenced items more prone to socially desirable responses (Miller & Simmering, 2023). In contrast, the hidden framing sample, which likely included only respondents with domain experience, exhibited a more evenly distributed CMV pattern. Importantly, the observed CMV did not result in CMB in either sample. To our knowledge, this framing-contingent CMV pattern has not been documented in prior methodological studies, implying that most standard data quality checks may miss potential impostor contamination.

Based on these findings, we recommend that researchers integrate the following practices into their survey design and data collection procedures (see Table 7).

Consideration	Description	Recommendation
Attention-trap questions have limited use for detecting impostors	Financially motivated and non-naive professional survey respondents know to look for careless-responding detectors (Meade & Craig, 2012; Peer et al., 2022), as failing them may result in denying of reward	Include attention-trap questions as part of a broader data quality assurance strategy, recognizing their limited utility for identifying impostors
Self-reported participant engagement measures are not useful for detecting impostors	Measures such as self-reported integrity, diligence, or interest are insufficient for identifying impostors or irrelevant respondents due to impression management, which may be further exacerbated when attention-trap questions are employed (Peer et al., 2022; Sharpe Wessling et al., 2017)	Elicit context-dependent expertise that is difficult to fake at multiple points in the survey. Compare responses for consistency to identify potential impostors
Many data quality metrics are insufficient to identify the presence of irrelevant respondents	Most data quality metrics (except for CMV) may not indicate lower data quality when the sample includes a mix of true target population and impostors. Impostors tend to avoid scrutiny from survey administrators that could jeopardize their rewards, resulting in superficially plausible data	Use CMV as a supportive diagnostic tool to assess whether the sample is likely to include ineligible respondents, as impostors may inflate positively valenced items due to low-effort responding
Hidden screeners complicate data collection and make it more cumbersome, but their use will likely result in a more relevant and valid sample	Including an in-study screener question set or conducting a separate screener study as part of a two-stage screening procedure is more expensive, time-consuming, and labor-intensive for the researcher, but it is likely to yield more relevant and higher-quality data (Chandler & Paolacci, 2017; Sharpe Wessling et al., 2017)	Calculate the required minimum sample size a priori based on the complexity of the model and assess budget sufficiency under a scenario in which up to 90% of the participants are screened out and compensated with a small amount (as required by the platform). When feasible, use a separate screener study as part of a two-stage process to minimize the influence of the higher total reward (Sharpe Wessling et al., 2017)

**Table 7. Recommendations for Survey Design and Data Collection in Paid Online Panels**

Depending on the context, using the explicit framing approach may still be appropriate, particularly when investigating a broad topic to which most individuals can reasonably respond. However, internal screeners provided by survey panel companies are not a panacea and should not be used as a justification for openly disclosing the survey topic. These screeners are often limited in scope and lack sufficient granularity. Moreover, participants may deliberately provide information that increases their chances of being selected for more surveys (e.g., claiming frequent use of artificial intelligence tools because it is perceived as a “hot topic”). Background information may also become outdated, such as the industry in which a participant works. Therefore, internal platform-level screeners are insufficient on their own, and the specific survey topic should generally not be disclosed upfront. Instead, hidden framing should generally be preferred, with respondents being thoughtfully screened into the survey using concealed eligibility criteria to ensure the collection of more relevant data. We see this as especially important when targeting users of specific IT environments (e.g., particular ERP or metaverse applications) or software-intensive and embedded systems where IT use is integral to daily work (e.g., vehicles and other cyber-physical systems).

### ***Limitations and future research***

Our study has certain limitations. First, our sample was collected from a single survey panel (Prolific), which has been assessed to have more honest respondents than other panels and is more likely to include non-naive professional survey takers (Peer et al., 2017, 2022). Future research should attempt to replicate the results of this study by utilizing data from other panels, as there may be variance in overall data quality provided by different panels depending on respondent composition and panel oversight.

Second, impostors in the explicit survey framing sample may still have operated work vehicles in contexts other than the one disclosed (e.g., long-haul trucking). Thus, they may have been able to provide credible answers inferred from their own context, leading to data that appears plausible and of high quality. The topic may have been sufficiently understandable for respondents with adjacent or no experience, enabling them to provide plausible answers. Therefore, it is important to replicate this study across multiple IS contexts. Future research could apply the same study design in a domain where “faking expertise” is more difficult, potentially yielding clearer distinctions between data provided by domain experts and non-experts.

Third, our study did not include specific measures aimed at identifying potential impostors. However, the substantial differences in screen-out rates between our samples suggest that more impostors are likely to be included when using an explicit survey framing approach. Future research could aim to create novel methods for detecting irrelevant respondents (Chandler & Paolacci, 2017; Kan & Drumme, 2018). Replicating the study in a context where the exact distribution of the initial participant pool is known or determined a priori would also be useful.

Fourth, the majority of the explicit framing sample was collected first to ensure that early respondents could be blocked from participating in the survey using the alternative framing condition. Given the relatively niche topic of the survey (non-road work vehicles users), it is possible that a substantial portion of the more active and engaged panelists fitting this target demographic were sampled initially, leaving fewer eligible respondents available for the hidden framing data collection. However, a small batch of the explicit framing sample was also collected after the hidden framing sample, and it followed a similar pattern. Nevertheless, future studies should consider modifying the study design to use an inverse or parallel sampling approach to avoid potentially “exhausting” the target sample population.

Fifth, the screener questions in the hidden framing condition were presented at the beginning of the study, with participants aware of the high reward associated with the full survey. This may have prompted respondents to consider their answers to the screening questions more carefully. Future studies should separate the initial screener survey (offering a substantially lower reward) and distribute the full survey to eligible participants later. This approach would allow researchers to assess whether differences emerge between the explicit framing and a more robust version of the hidden framing (Sharpe Wessling et al., 2017).

### **Acknowledgements**

This research was supported by the Business Finland project Decarbonization of Mobile Machine Systems (Decarbo) (7942/31/2022) and the Evil-AI project (funded by Jane and Aatos Erkko Foundation).

## References

- Arndt, A. D., Ford, J. B., Babin, B. J., & Luong, V. (2022). Collecting samples from online services: How to use screeners to improve data quality. *International Journal of Research in Marketing*, 39(1), 117–133. <https://doi.org/10.1016/j.ijresmar.2021.05.001>
- 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. <https://doi.org/10.1007/s11162-011-9251-2>
- Belliveau, J., & Yakovenko, I. (2022). Evaluating and improving the quality of survey data from panel and crowd-sourced samples: A practical guide for psychological research. *Experimental and Clinical Psychopharmacology*, 30(4), 400–408. <https://doi.org/10.1037/phaaaa00564>
- Berinsky, A. J., Margolis, M. F., & Sances, M. W. (2014). Separating the Shirkers from the Workers? Making Sure Respondents Pay Attention on Self-Administered Surveys. *American Journal of Political Science*, 58(3), 739–753. <https://doi.org/10.1111/ajps.12081>
- 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. <https://doi.org/10.1016/j.chb.2013.05.009>
- Chandler, J. J., & Paolacci, G. (2017). Lie for a Dime: When Most Prescreening Responses Are Honest but Most Study Participants Are Impostors. *Social Psychological and Personality Science*, 8(5), 500–508. <https://doi.org/10.1177/1948550617698203>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Clifford, S., & Jerit, J. (2015). Do Attempts to Improve Respondent Attention Increase Social Desirability Bias? *Public Opinion Quarterly*, 79(3), 790–802. <https://doi.org/10.1093/poq/nfv027>
- Compeau, D., Marcolin, B., Kelley, H., & Higgins, C. (2012). Generalizability of Information Systems Research Using Student Subjects—A Reflection on Our Practices and Recommendations for Future Research. *Information Systems Research*, 23(4), 1093–1109. <https://doi.org/10.1287/isre.1120.0423>
- Fang, Y.-H., & Li, C.-Y. (2022). Does the sharing economy change conventional consumption modes? *International Journal of Information Management*, 67, 102552. <https://doi.org/10.1016/j.ijinfomgt.2022.102552>
- Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. *Journal of Marketing Research*, 18(1), 39–50. <https://doi.org/10.2307/3151312>
- Gaskin, J., James, M., & Lim, J. (2019). Master Validity Tool, AMOS Plugin. Gaskination's StatWiki.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis: Pearson New International Edition* (7th edition). Pearson.
- Hays, R. D., Liu, H., & Kapteyn, A. (2015). Use of Internet panels to conduct surveys. *Behavior Research Methods*, 47(3), 685–690. <https://doi.org/10.3758/s13428-015-0617-9>
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Hyman, M. R., & Sierra, J. J. (2012). Adjusting Self-Reported Attitudinal Data for Mischievous Respondents. *International Journal of Market Research*, 54(1), 129–145. <https://doi.org/10.2501/IJMR-54-1-129-145>
- Kan, I. P., & Drumrey, A. B. (2018). Do imposters threaten data quality? An examination of worker misrepresentation and downstream consequences in Amazon's Mechanical Turk workforce. *Computers in Human Behavior*, 83, 243–253. <https://doi.org/10.1016/j.chb.2018.02.005>
- Keusch, F., Batinic, B., & Mayerhofer, W. (2014). Motives for joining nonprobability online panels and their association with survey participation behavior. In M. Callegaro, R. Baker, J. Bethlehem, A. S. Göritz, J. A. Krosnick, & P. J. Lavrakas (Eds.), *Online Panel Research* (1st ed., pp. 171–191). Wiley. <https://doi.org/10.1002/9781118763520.ch8>
- Kim, H.-W., & Kankanhalli, A. (2009). Investigating user resistance to information systems implementation: A status quo bias perspective. *MIS Quarterly*, 33(3), 567–582. <https://doi.org/10.2307/20650309>

- Kwak, D.-H., Holtkamp, P., & Kim, S. S. (2019). Measuring and Controlling Social Desirability Bias: Applications in Information Systems Research. *Journal of the Association for Information Systems*, 20(4), 317–345. <https://doi.org/10.17705/1jais.00537>
- Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence Testing for Psychological Research: A Tutorial. *Advances in Methods and Practices in Psychological Science*, 1(2), 259–269. <https://doi.org/10.1177/2515245918770963>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>
- Miller, B. K., & Simmering, M. J. (2023). Attitude Toward the Color Blue: An Ideal Marker Variable. *Organizational Research Methods*, 26(3), 409–440. <https://doi.org/10.1177/10944281221075361>
- Peer, E., Brandimarte, L., Samat, S., & Acquisti, A. (2017). Beyond the Turk: Alternative platforms for crowdsourcing behavioral research. *Journal of Experimental Social Psychology*, 70, 153–163. <https://doi.org/10.1016/j.jesp.2017.01.006>
- Peer, E., Rothschild, D., Gordon, A., Evernden, Z., & Damer, E. (2022). Data quality of platforms and panels for online behavioral research. *Behavior Research Methods*, 54(4), 1643–1662. <https://doi.org/10.3758/s13428-021-01694-3>
- Pickering, D., & Blaszczynski, A. (2021). Paid online convenience samples in gambling studies: Questionable data quality. *International Gambling Studies*, 21(3), 516–536. <https://doi.org/10.1080/14459795.2021.1884735>
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of Method Bias in Social Science Research and Recommendations on How to Control It. *Annual Review of Psychology*, 63(1), 539–569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- Podsakoff, P. M., Podsakoff, N. P., Williams, L. J., Huang, C., & Yang, J. (2024). Common Method Bias: It's Bad, It's Complex, It's Widespread, and It's Not Easy to Fix. *Annual Review of Organizational Psychology and Organizational Behavior*, 11, 17–61. <https://doi.org/10.1146/annurev-orgpsych-110721-040030>
- Polites, G. L., & Karahanna, E. (2012). Shackled to the Status Quo: The Inhibiting Effects of Incumbent System Habit, Switching Costs, and Inertia on New System Acceptance. *MIS Quarterly*, 36(1), 21–42. <https://doi.org/10.2307/41410404>
- Protzko, J., Zedelius, C. M., & Schooler, J. W. (2019). Rushing to Appear Virtuous: Time Pressure Increases Socially Desirable Responding. *Psychological Science*, 30(11), 1584–1591. <https://doi.org/10.1177/0956797619867939>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the Fit of Structural Equation Models: Tests of Significance and Descriptive Goodness-of-Fit Measures. *Methods of Psychological Research Online*, 8(2), 23–74.
- Schlenker, B. R. (2008). Integrity and Character: Implications of Principled and Expedient Ethical Ideologies. *Journal of Social and Clinical Psychology*, 27(10), 1078–1125. <https://doi.org/10.1521/jscp.2008.27.10.1078>
- Schmalfuß, F., Mühl, K., & Krems, J. F. (2017). Direct experience with battery electric vehicles (BEVs) matters when evaluating vehicle attributes, attitude and purchase intention. *Transportation Research Part F: Traffic Psychology and Behaviour*, 46, 47–69. <https://doi.org/10.1016/j.trf.2017.01.004>
- Sharpe Wessling, K., Huber, J., & Netzer, O. (2017). MTurk Character Misrepresentation: Assessment and Solutions. *Journal of Consumer Research*, 44(1), 211–230. <https://doi.org/10.1093/jcr/ucx053>
- Stefkovics, Á., Eichhorst, A., Skinnion, D., & Harrison, C. H. (2024). Are We Becoming More Transparent? Survey Reporting Trends in Top Journals of Social Sciences. *International Journal of Public Opinion Research*, 36(2), edae013. <https://doi.org/10.1093/ijpor/edae013>
- Tarafdar, M., Tu, Q., Ragu-Nathan, B. S., & Ragu-Nathan, T. S. (2007). The Impact of Technostress on Role Stress and Productivity. *Journal of Management Information Systems*, 24(1), 301–328. <https://doi.org/10.2753/MIS0742-1222240109>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425–478. <https://doi.org/10.2307/30036540>
- Williams, L. J., Hartman, N., & Cavazotte, F. (2010). Method variance and marker variables: A review and comprehensive cfa marker technique. *Organizational Research Methods*, 13(3), 477–514. <https://doi.org/10.1177/1094428110366036>