Asking People

The current state of human research, 2026

A FishDog white paper

Author: Andreas Duess, CEO

July 2026

Contents

Executive summary

The global insights industry generated $153 billion in revenue in 2024. Most of that spend runs through one instrument: a questionnaire, put to people who agreed to answer it. This paper reviews the published record on the current condition of that instrument. Its findings, in the order the paper presents them:

Participation. Pew Research Center's telephone response rates fell from 36 percent in 1997 to 6 percent in 2018. Its probability-based online panel reports a 92 percent wave response rate and a 3 percent cumulative response rate, counting the full path from sampled population to delivered answer.

Who answers. The average online panellist takes nearly 11 surveys per week, and roughly one in five takes more than 25. In one platform sample, 40 percent of participants had taken seven or more surveys in the previous 24 hours. Fourteen percent of US adults are current paid survey panellists.

What answers are made of. Thirty-four percent of platform participants report using large language models on open-ended questions, and the resulting text is measurably more homogeneous than human writing. An autonomous AI respondent documented in PNAS in 2025 passed 99.8 percent of attention checks and evaded every detection method in current use, at a cost of about five cents per completed interview.

Contamination. A six-source test flagged 29 percent of respondents as suspicious or fraudulent. A 2026 NORC literature review places typical industry fraud at 15 to 30 percent. Client rejections of delivered sample rose about 300 percent over three years.

Detection. Of the respondents Pew identified as bogus, 84 percent passed an attention-check trap question and 87 percent passed a speeding check. Opt-in samples average 5.8 percentage points of error against government benchmarks; probability panels average 2.6.

The instrument. Question wording and order move results by up to 25 percentage points on identical populations. Stated purchase intent predicts sales under narrow, documented conditions, weakest for new products over long horizons.

The paper concludes that survey-based research is a measurement instrument with quantified error from four sources: who agreed to be asked, who did the answering, how the questions were worded and ordered, and what the respondent used to compose the reply. The output of such an instrument is a reading rather than a fact. The paper closes with six questions a buyer can put to any research supplier, drawn from the findings above.

Scope

In 2024 the global insights industry generated $153 billion in revenue, up from $142 billion the year before. Sixty-four percent of research spend is quantitative, and nearly all recent growth in quantitative spend has come from online surveys. The instrument at the centre of that spend is the questionnaire, fielded to people who agreed to answer it.

This paper describes the current condition of that instrument: participation, respondent behaviour, contamination levels, the performance of standard quality controls, and the measurement properties of the questionnaire itself. Full citations are given at the end of the paper.

Part I: Participation and sample composition

1. Response rates

Pew Research Center's telephone surveys had a 36 percent response rate in 1997. By 2012 the rate was 9 percent. It held there for several years, then resumed falling: 7 percent in 2017, 6 percent in 2018.

Online panels report response rates differently. Pew's American Trends Panel, a probability-based online panel, reports a 92 percent response rate for a wave fielded November 17 to 30, 2025: 10,357 of 11,258 sampled panellists responded.

The same methodology statement reports the cumulative response rate, which accounts for nonresponse at recruitment and attrition since. It is 3 percent.

Both figures describe the same panel. The first counts people who answered once they were already members. The second counts the full path from sampled population to delivered response.

2. Respondent frequency and panel overlap

Morning Consult surveyed 4,420 of its own panellists in 2022. The average respondent sourced from online panels takes nearly 11 surveys per week, and roughly 20 percent take more than 25 per week.

A 2025 study in *Sociological Methods & Research* by Simone Zhang, Janet Xu, and AJ Alvero drew on roughly 800 participants from a major online research platform. Forty percent had taken seven or more surveys in the previous 24 hours.

CivicScience, surveying more than one million US adults, found that 14 percent are current paid survey panellists and a further 12 percent are former panellists.

Panel membership also overlaps across suppliers. In a Morning Consult analysis of more than 300,000 US responses, 40 percent of repeat respondents from one panel had previously entered the ecosystem through a different source.

3. The composition of answers

In the Zhang, Xu, and Alvero study, 34 percent of participants reported using large language models to help them answer open-ended survey questions.

The authors compared human-written responses from three pre-ChatGPT studies against machine-generated text. From the abstract: "LLM responses are more homogeneous and positive, particularly when they describe social groups in sensitive questions. These

homogenization patterns may mask important underlying social variation in attitudes and beliefs among human subjects, raising concerns about data validity."

The pattern is not confined to consumer panels. A 2023 EPFL study estimated that 33 to 46 percent of Amazon Mechanical Turk crowd workers used LLMs to complete a text summarisation task.

A portion of survey responses is not composed by a panellist at all. Sean Westwood of Dartmouth built an autonomous AI respondent from a prompt of roughly 500 words. As published in *PNAS* in 2025, it passed 99.8 percent of attention checks, made zero errors on logic puzzles, and evaded every detection method currently in use. It operated from Russian, Mandarin, and Korean language settings while producing fluent English responses, and tailored its answers to whatever demographic profile it was assigned. A completed interview cost about five cents, against a typical human incentive of $1.50. Westwood calculated that adding between 10 and 52 such responses to any of the seven major national polls before the 2024 US election would have changed the predicted outcome.

4. Contamination estimates

Estimates of fraudulent, duplicate, or automated traffic in online sample are available from several independent sources.

Rep Data ran the same census-balanced survey across six online sample sources in 2025 and flagged an average of 29 percent of respondents as suspicious or fraudulent. The same study found hyperactive respondents attempting 70 or more surveys per day; one attempted more than 1,100 in a single day.

Rep Data's fraud-detection arm scanned 4.1 billion survey attempts between January and August 2025 and flagged roughly one third as fraudulent, with a further 27 percent classified as inattentive.

A NORC literature review published in April 2026 places typical industry fraud rates at 15 to 30 percent, "as high as 45 percent on some survey platforms," and documents academic case studies in which usable responses fell from roughly 75 percent of completes to 10 percent in recent years.

Kantar reports that reconciliation rates, the share of delivered sample rejected by clients after fieldwork, increased by around 300 percent over three years, with some clients rejecting up to 40 percent of data post-field.

Contamination predates generative AI. In 2018, researchers auditing Amazon Mechanical Turk found that about 20 percent of respondents that summer were connecting through server farms or non-US IP addresses. Follow-up work traced the traffic not to bots but to organised human respondents misrepresenting their location, most of whom failed basic English screeners while passing every automated bot check. A four-wave study across the same period found significant

increases in failed validity checks, measurable declines in the reliability of a standard personality instrument, and failures to replicate well-established findings.

For context on the wider environment, Imperva's 2024 Bad Bot Report, cited in the NORC review, estimates that bots account for nearly half of all internet traffic.

5. Compensation

A 2018 ACM study instrumented 2,676 workers performing 3.8 million tasks on Amazon Mechanical Turk. Median hourly earnings were approximately $2. Four percent of workers earned more than the US federal minimum wage of $7.25. Requesters paid more than $11 per hour on average, but lower-paying requesters posted most of the available work.

Consumer panel rates sit in the same range. One 2025 tested comparison measured effective earnings of $2.85 per hour on Swagbucks and $4.20 per hour on Survey Junkie.

Brosnan, Kemperman, and Dolnicar, in the *International Journal of Market Research*, quantified what drives panellists to accept a survey: the incentive accounts for 29 percent of the participation decision, and speed of completion ranks second at 14 percent. They classify 15 percent of panel members as "mercenary responders," motivated primarily by the payment. Across these platforms, payment is conditional on completing the survey; no element of the standard incentive structure is conditional on the quality of the answers.

Part II: Quality controls and industry response

6. Detection performance

The standard quality controls are attention checks, speeding thresholds, and consistency tests. Pew Research Center measured their performance.

Across more than 60,000 interviews in 2019, bogus respondents made up 4 to 7 percent of completes in opt-in panels and 1 percent in address-recruited probability panels. Of the bogus respondents, 84 percent passed an attention-check trap question, 87 percent passed a speeding check, and 76 percent passed both.

CloudResearch, vetting more than 127,000 respondents in 2021, found that about 30 percent of panel respondents routinely fail simple attention checks. Among respondents its screening system rejected, 51 percent claimed to be petroleum engineers. Of 253 people claiming to live with an artificial limb, 99 percent were judged to be lying.

Screener misrepresentation is documented in the academic record as well. Chandler and Paolacci showed in 2017 that when studies screen for rare populations, a large share of the participants who qualify are impostors: most prescreening answers are honest, but the dishonest minority concentrates where eligibility pays.

Pew applied the same logic to opt-in samples in 2024 with a question about holding a licence to operate a nuclear submarine. Twelve percent of opt-in respondents under 30 claimed one. The true rate is effectively zero.

7. Industry initiatives

The Global Data Quality initiative launched in 2023, bringing together ESOMAR, the Insights Association, the MRS, and nine other industry bodies around fraud tracking, incentive practice, and shared standards. Imperium launched a data quality certification in 2020, with Cint and Dynata among the first certified; Imperium's own research put the share of respondents who are duplicates, fraudsters, or bots at up to 38 percent.

Adoption of countermeasures is now widespread. The 2026 GRIT Insights Practice report finds fraud-detection tools in regular use by 70 to 88 percent of firms across every industry segment. None of the methods in current use detected Westwood's autonomous respondent.

8. Accuracy by sample type

Measured accuracy differs by sample type. Pew's 2023 benchmarking study fielded identical questions to opt-in samples and probability-based panels, then compared both against 28 government benchmarks. Opt-in samples averaged 5.8 percentage points of error per estimate. Probability panels averaged 2.6.

For adults under 30, opt-in error rose to 11.2 points against 3.6. For Hispanic adults, 10.8 against 3.6. In a December 2023 opt-in poll, 20 percent of under-30 respondents agreed that the Holocaust is a myth; on Pew's probability panel the figure was 3 percent.

At the project level, US agencies quote roughly $5,000 to $15,000 for a 400-response online survey, $7,000 to $20,000 per focus group, and $25,000 to $65,000 for a custom qualitative or quantitative project, on timelines of two days to four weeks for online fieldwork and three to five weeks for qualitative programs.

Part III: Measurement properties

9. Question wording and order

Responses are also sensitive to the design of the instrument itself. Pew Research Center publishes its own question-design experiments, each fielded to randomly assigned groups, so the differences are attributable to the question rather than to the people answering it.

Experiment	Version A	Version B	Difference
Military action in Iraq, January 2003	68 percent favoured "taking military action in Iraq to end Saddam Hussein's rule"	43 percent favoured the same action when the question added "even if it meant that U.S. forces might suffer thousands of casualties"	25 points
Most important issue, 2008	58 percent named the economy when it appeared as a listed option	35 percent named the economy when the question was open-ended	23 points
National satisfaction, December 2008	88 percent dissatisfied with the country's direction when asked after a presidential approval question	78 percent dissatisfied without that preceding question	10 points
Legal agreements for same-sex couples, October 2003	45 percent in favour when asked after a question about marriage	37 percent in favour without that preceding question	8 points
Assisted dying, 2005	51 percent favoured "the means to end their lives"	44 percent favoured "assist terminally ill patients in committing suicide"	7 points

The underlying pattern was established by Schuman and Presser across more than 200 experiments in 34 surveys, published in 1981. Malhotra showed in 2008 that the fastest respondents are the most vulnerable to response-order effects. Galesic and Bosnjak showed in 2009 that longer questionnaires reduce both participation and the quality of answers given later in the instrument, where responses come faster, skip more items, and contain less.

Respondent behaviour interacts with instrument design. In a probability-based Dutch panel, respondents answered a mean of 15 out of 54 questions faster than they could plausibly have

read them. On grid questions, 31 percent of those speeders straight-lined, against under 1 percent of other respondents. Longer-tenured panellists sped more.

10. Stated intent and observed behaviour

The relationship between what people tell a survey and what they subsequently do is conditional, and the conditions are documented.

Morwitz, Steckel, and Gupta, in the *International Journal of Forecasting*, established when purchase intentions predict sales: better for existing products than new ones, for short time horizons than long ones, and for specific brands than whole categories. Prediction is weakest for new products over long horizons, the conditions under which pre-launch research is most often commissioned.

The same gap appears in sustainability research. In *Harvard Business Review* in 2019: 65 percent of consumers said they want to buy from purpose-driven, sustainable brands. About 26 percent did.

11. The industry's own statistics

The most widely quoted statistic in market research holds that 80 to 95 percent of new products fail. The 95 percent figure, commonly attributed to Clayton Christensen, appears in no publication of his and circulates without a primary source. Nielsen's 2014 figure, 85 percent of new CPG products failing, specifies neither a timeframe nor a definition of failure. The peer-reviewed literature, reviewing studies that cover more than 1,000 business units across ten industries, puts the actual new product failure rate at approximately 40 percent and attributes the higher figures to repetition and self-interest. The paper making that case is titled "New Product Failure Rates: Influence of *Argumentum ad Populum* and Self-Interest."

Replication has been measured directly. A 2022 audit in *Frontiers in Communication* re-ran ten influential sensory marketing effects with a sample of 823 attentive respondents. Two of the ten replicated.

12. Conclusion

Survey-based research carries error from four sources: who agreed to be asked, who did the answering, how the questions were worded and ordered, and what the respondent used to compose the reply.

Each of these is now quantified in the published record. Cumulative response paths run at 3 percent. Between a quarter and a third of raw sample is flagged before or after fieldwork. A third of open-ended answers involve a language model. Question wording moves results by as much as 25 points. Stated intent predicts behaviour under narrow, documented conditions.

Instruments with known error remain usable, and measurement disciplines routinely work with them. Their output is a reading rather than a fact: produced by a specific sample, through a specific questionnaire, under specific economics, at a specific point in the condition of the panel ecosystem.

The record supports a set of questions that apply to any survey-based research, from any supplier:

- Who answered, and through what recruitment path? What is the cumulative response rate, not the wave rate?
- How many surveys had these respondents taken that week?
- What share of raw completes was removed, by what checks, and what do those checks demonstrably miss?
- What were respondents paid, and for what behaviour does the payment structure select?
- How sensitive is the headline number to wording and order? Was it tested?
- Where has the method been benchmarked against an outcome that occurred: a sales figure, an election, a re-run of a prior study?

Sources

Bispo Júnior, J. P. (2022). Social desirability bias in qualitative health research. *Revista de Saúde Pública*, 56, 101.

Brosnan, K., Kemperman, A., & Dolnicar, S. (2021). Maximizing participation from online survey panel members. *International Journal of Market Research*, 63(4), 416–435.

Castellion, G., & Markham, S. K. (2013). New product failure rates: Influence of argumentum ad populum and self-interest. *Journal of Product Innovation Management*, 30(5), 976–979.

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.

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science*, 11(4), 464–473.

CivicScience. Beware the paid survey panelist. CivicScience.

CloudResearch (Moss, A.) (2021, October 20). The truth about online data quality.

Drive Research (2026). How much does market research cost? / How long does market research take?

ESOMAR / Research World: Palacio, X. (2024, November 20). Drivers of our $142bn insights industry. Ajaluni, L. (2025, November 4). Inside the $153bn insights industry. Palacio, X., & Ajaluni, L. (2024, April 30). The ongoing transformation of market research methods.

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.

Global Data Quality initiative (2023). globaldataquality.org.

GreenBook (2026). GRIT Insights Practice Report, 16th edition.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). A data-driven analysis of workers' earnings on Amazon Mechanical Turk. *ACM CHI Conference on Human Factors in Computing Systems*. arXiv:1712.05796.

Imperium (2020, September 21). Imperium announces pioneering data quality certification program. PR Newswire.

Kantar. Is panel fraud the new ad fraud? Kantar.com.

Kennedy, C., & Hartig, H. (2019, February 27). Response rates in telephone surveys have resumed their decline. Pew Research Center.

Kennedy, C., Hatley, N., Lau, A., Mercer, A., Keeter, S., Ferno, J., & Asare-Marfo, D. (2020, February 18). Assessing the risks to online polls from bogus respondents. Pew Research Center.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. G. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods*, 8(4), 614–629.

Malhotra, N. (2008). Completion time and response order effects in web surveys. *Public Opinion Quarterly*, 72(5), 914–934.

Mercer, A., & Lau, A. (2023, September 7). Comparing two types of online survey samples. Pew Research Center.

Mercer, A., Kennedy, C., & Keeter, S. (2024, March 5). Online opt-in polls can produce misleading results, especially for young people and Hispanic adults. Pew Research Center.

Morning Consult (Podkul, A., & Martherus, J.) (2023, May 12). Collecting unique, high-quality respondents in today's online sample marketplace.

Morwitz, V. G., Steckel, J. H., & Gupta, A. (2007). When do purchase intentions predict sales? *International Journal of Forecasting*, 23(3), 347–364.

Motoki, K., & Iseki, S. (2022). Evaluating replicability of ten influential research on sensory marketing. *Frontiers in Communication*, 7, 1048896.

NORC at the University of Chicago (2026, April). Fraudulent respondents and bots in nonprobability surveys: A literature review. Research brief.

Pew Research Center (2012, May 15). Assessing the representativeness of public opinion surveys.

Pew Research Center (2017, May 15). What low response rates mean for telephone surveys.

Pew Research Center (2026, June 10). Typology 2026, Appendix A: Survey methodology.

Pew Research Center. Writing survey questions. Methods.

Rep Data (2025, June 2). What new research-on-research says about fraud in survey data. Rep Data (Snell, S.) (2025). State of Survey Fraud 2025, as reported by GreenBook.

Schuman, H., & Presser, S. (1981). *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context.* Academic Press.

TurkPrime / CloudResearch (Moss, A., & Litman, L.) (2018, September 18). After the bot scare: Understanding what's been happening with data collection on MTurk.

Veselovsky, V., Horta Ribeiro, M., & West, R. (2023). Artificial artificial artificial intelligence: Crowd workers widely use large language models for text production tasks. arXiv:2306.07899.

Westwood, S. J. (2025). The potential existential threat of large language models to online survey research. *Proceedings of the National Academy of Sciences*, 122(47), e2518075122.

White, K., Hardisty, D. J., & Habib, R. (2019, July–August). The elusive green consumer. *Harvard Business Review*.

Zhang, C., & Conrad, F. G. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods*, 8(2), 127–135.

Zhang, S., Xu, J., & Alvero, A. (2025). Generative AI meets open-ended survey responses: Research participant use of AI and homogenization. *Sociological Methods & Research*, 54(3), 1197–1242.