walr.

Whitepaper · 2026

Data quality

Optimizing for understanding.

Why question design is the next frontier in data quality.

Survey 1 · Standard

Measurement

Survey 2 · Conversational

Understanding

Written by

Staci Kinney, Walr

An A/B test of survey design

2,000 people · One survey · Two experiences

Contents

What's inside

We've been optimizing for measurement. We should be optimizing for understanding.

We've spent years getting better at one question: is this a real, engaged human? How panels are built and how we use technology has paved the way for huge progress here. But there's a second question we ask far less often. Once we know we're talking to a real person, how hard are we working to talk to them like one?

To find out, we ran an A/B test on survey design. We gave 2,000 people the same survey on AI and autonomous driving, changing only how the questions were worded. The conversational version improved engagement and clarity, and it didn't cost us rigor.

Finding 01

Same signal

The data told the same story. Rank order held across nearly every question.

Finding 02

Better engagement

Junk open-ends fell from 19% to 8%, same fraud controls. People stayed longer and wrote more.

Finding 03

Greater clarity

The meaning behind each answer was easier to read. Less decoding, more deciding.

Have we optimized the survey at the expense of the respondent?

Somewhere along the way, surveys changed. They started as tools to understand humans. They became exercises in satisfying research best practice. Most ad hoc research is custom, yet we keep forcing new questions into old frameworks because they fit established metrics. The result? We spend more time protecting the measurement than understanding the human.

It's worth asking what would happen if we flipped that. Start by understanding the consumer as intuitively as possible. Then build the metric around that understanding - not the other way around.

We've solved the “who.”

Before you can understand a human, you have to be sure you're talking to one. No metric, methodology, or questionnaire survives poor respondent quality.

This is where the industry has made real progress. At Walr, we have a combination of our own proprietary technology and third party tools to stop fraud and bad actors before they ever reach a survey. It’s important work, and we will continue to invest in this area. However, it's time to look at the next lever we can pull.

There are many levers to improve data quality. But there is one we overlook more than most.

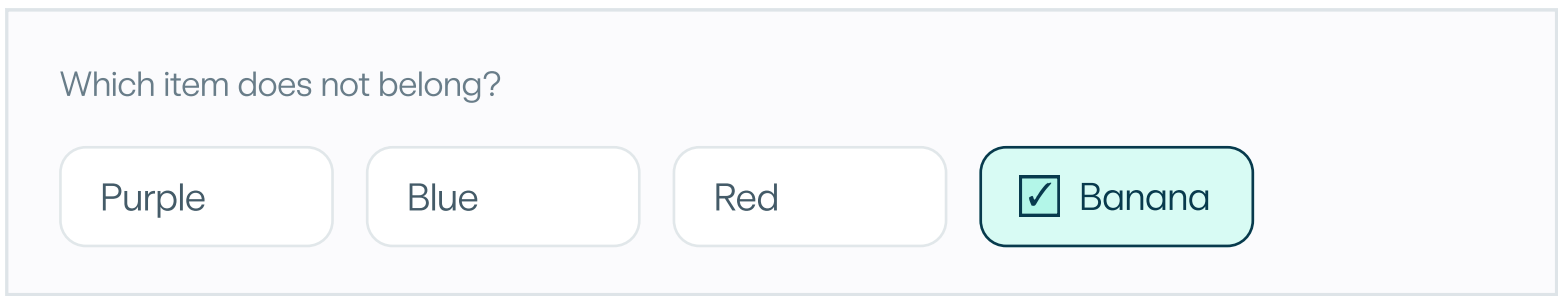
How we ask questions.

A good principle that holds up everywhere: say what you mean, and mean what you say. The same applies to market research. Yes, avoid double-barrelled questions. Yes, follow methodological best practice. But those are guardrails, not the destination. The real job is to understand the business question, then ask people about it in the clearest, most natural way possible.

A perfectly balanced 5-point scale is methodologically sound. But if the scale isn't relevant to the question, or the question isn't relevant to the decision, we've just collected cleaner data about the wrong thing.

The attention-check moment

We spend enormous effort screening for real, attentive respondents. Then we ask them to identify a giraffe, a hot air balloon, or a baseball field. Or to spot which item doesn't belong:



A verified human just spent their attention here. Does this still serve us?

These checks have a purpose. But they're also how it's always been done, and that's worth questioning. Every one of these moments reminds a respondent they're taking a survey, not having a conversation. And the moment that happens, engagement starts to slip away.

The challenge isn't only finding high-quality respondents. It's keeping high-quality respondents engaged long enough to give high-quality answers.

05 The hypothesis

The more human the experience, the richer the understanding.

When surveys feel like tasks, respondents give us answers. When surveys feel natural, they give us clarity. They don't just tell us what they think. They help us understand why. And that's what clients are really looking for: not data points, but the context, nuance, and human truth behind them.

One survey.
Two experiences.

We gave 2,000 people the same survey on AI and autonomous driving. 1,000 people completed a traditional version, using standard market-research language and scales. 1,000 people completed a conversational version, using natural language and a more human tone.

Same objectives. Same nationally representative sample. Same topics. The only difference was the respondent experience.

A Standard · Scale-based

What is the main reason you would be hesitant to trust a fully driverless vehicle?

Please select one answer

- ☐ Lack of accountability if something goes wrong
- ☐ I simply prefer having a human in control
- ☑ Privacy - who has access to my journey data?
- ☐ I don’t fully understand how the technology works
- ☐ Concerns about safety in unexpected situations
- ☐ I don’t have hesitations - I would trust it

B Conversational · Human

What's the main thing holding you back from fully trusting a driverless car?

Please select one answer

- ☐ What happens when something totally unexpected occurs?
- ☐ The data angle - someone knowing everywhere I go
- ☑ Who's responsible if it crashes?
- ☐ I don't really get how it works, and that unsettles me
- ☐ I just like being in control. Simple as.
- ☐ Nothing - I'd hop right in

A standard survey and a conversational survey, side by side. Survey 1 is optimized for measurement; Survey 2 for human expression.

Same signal.
Better experience.

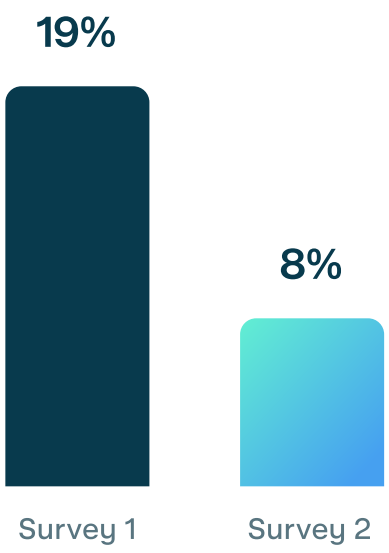
Across nearly every question, the two versions tracked each other closely. **Rank order held:** cost led on frustrations, traffic followed. **Trust in AI tasks ran the same way:** GPS navigation highest, a full door-to-door trip lowest. **Hesitation came down to the same factors:** a preference for human control, and safety in unexpected situations.

Skepticism also ran deep in both groups. The majority were reluctant to use a driverless taxi today, and safety perception skewed negative. Wariness, not enthusiasm, was the primary position.

The fact the signal was consistent, matters. It gives us permission to talk human-to-human without sacrificing methodological rigor.

Engagement pulled ahead.

This is where the conversational version pulled ahead. In the traditional survey, 19% of open-end responses to a question on trust were junk: gibberish, off-topic, or otherwise unusable. In the conversational version, that dropped to 8%.



Junk On A Question About Trust

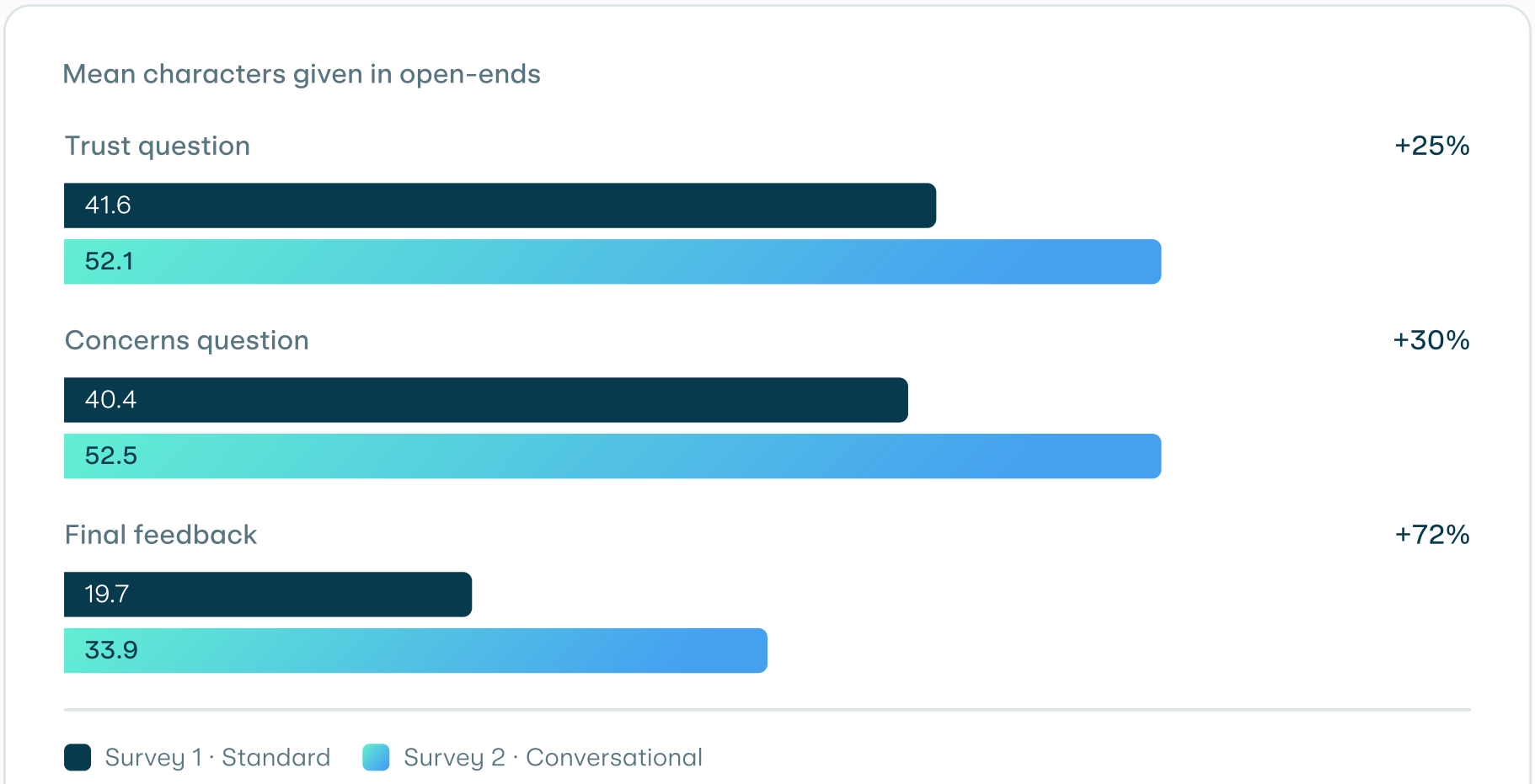
−58%

relative drop in junk open-ends, with identical fraud controls in both cells.

More to say, right to the end.

Here's the part that matters most. Both groups went through the exact same quality controls. These were real, verified respondents on both sides. This wasn't a panel-quality difference. It was an engagement difference.

We saw it in the writing. On the question relating to trust, conversational respondents typed 25% more characters than the traditional group. By the final feedback question, exactly where survey fatigue normally compresses answers most, the gap widened to 72%. People also chose to answer more often, on every open-end question.



The data also showed how length of interview (LOI) doesn't necessarily equal poor quality, with the conversational survey leading to a higher time spent in the survey. This signals it may be best viewed as a quality indicator, not a fixed metric of success.

Time In Survey (Mean)

7.18 → 9.54

minutes. Engagement, not abandonment.

Would Take Another Survey

46% → 53%

a 7-point lift in repeat intent.

The story got easier to read.

The conversational survey didn't change the story. It made the story easier to understand. Traditional scales often need translating. A researcher has to work out what a “4” or a “5” actually means for the consumer, and then for the business. Conversational attributes and scales reduce that step. The emotional intent behind each answer is clearer.

Our value isn't in explaining what someone meant by picking a number. It's in understanding the human behind that answer, and helping clients make better decisions because of it. Conversational design gets us there faster. Less time decoding what respondents said. More time explaining why it matters.

+7 pts more people said they'd take another survey after the conversational version (46% → 53%).

These weren't better respondents. They were good respondents having a better experience.

You don't need to reinvent the wheel. Just adopt some new habits.

Conversational design isn't a rewrite of your questionnaire. It's a shift in who the question is written for. Here's where we would advise starting.

01

Start with the business question, then work backwards.

Write down the decision the client needs to make. Then ask about it in the clearest, most natural way you can.

02

Open like a conversation, not a test.

Same intent. One sounds like a form. The other sounds like a person who's curious about your answer.

03

Write answer options the way people actually talk.

“Honestly nothing” earns more truthful answers than a flat “None of the above.” People recognize themselves in the options.

04

Keep the scale. Lose the translation step.

Where the numbers aren't serving the decision, anchor them in language the respondent and the client both understand instantly.

05

Audit your attention checks.

Keep the ones that protect data quality, question the ones that only protect habit.

The next frontier in data quality isn't a filter.

It's the question itself.

The frontier is moving. For years, better data quality meant keeping the wrong people out. Increasingly, it means getting more from the right ones.

That's less a screening problem, more a writing one. The teams who pull ahead will be the ones who treat each question as something a respondent actually chose to answer, and write it that way.

The patterns on the previous pages are where any team can start.

So pick one study. Rewrite a single question the way you'd ask it out loud. Then read what comes back, and judge for yourself.

If you want to talk about what this could look like in your own studies.

Let's talk →

A/B test of survey design. The same AI and autonomous-driving survey was served to two groups. Survey 1 (control) used standard, neutral wording and scales. Survey 2 (experimental) used conversational wording, answer choices written the way a real person speaks.

Bases: Survey 1 n = 1,056 completes, Survey 2 n = 1,024 completes. Both cells passed identical fraud and quality controls (Walr Gatekeeper, incorporating Sentry and Verisoul).

Satisfaction. Identical 1-5 anchors in both cells.

Metric	Survey 1	Survey 2	Change
Enjoyment CSAT (1-5 mean)	3.66	3.74	+0.08
Repeat-intent CSAT (1-5 mean)	4.03	4.20	+0.17
Repeat-intent top-box	45.8%	52.8%	+7.0 pts

Open-end engagement. Mean characters written, by question.

Open-end question	Survey 1	Survey 2	Change
Trust	41.6	52.1	+25.3%
Concerns	40.4	52.5	+29.9%
Final feedback	19.7	33.9	+72.0%

Open-end quality. Scored 1-5 (1 = junk, 5 = detailed), applied identically to both cells.

Open-end question	Survey 1 mean	Survey 2 mean	Change
Trust	2.41	2.68	+11.2%
Concerns	2.34	2.55	+9.0%
Final feedback	1.83	2.31	+26.5%

On the trust question, Score 1 (junk) responses fell from 19.0% in Survey 1 to 8.4% in Survey 2. Mean length of interview rose from 7.18 minutes (Survey 1) to 9.54 minutes (Survey 2).