SURVEYS · QUESTIONNAIRE DESIGN

Principles of Questionnaire Design: Clarity, Brevity and Neutrality

A questionnaire is a measuring instrument, and small flaws in its wording produce large errors in its results. This article sets out the three core principles of good question writing, the research behind them, and practical rules for applying them.

ModQu research article · 2,552 words · 26 references (APA 7th edition)

A survey is only as good as its questions. Careful sampling, high response rates and sophisticated analysis cannot rescue data collected with questions that respondents misunderstand, cannot answer, or are nudged to answer in a particular way. Yet questionnaires are often written quickly, by people who know the subject well and assume that everyone else will read the questions the same way they do.

Decades of experimental research show that this assumption is often wrong. Small changes in wording, order and format can shift answers by many percentage points (Schuman & Presser, 1981). The response is not to treat question writing as a matter of taste, but to follow principles that have been tested repeatedly. Three of these principles underpin almost all specific advice: **clarity**, **brevity** and **neutrality**. This article explains each in turn, describes the evidence behind them and offers practical rules for applying them.

Why question wording matters

Answering a survey question is a demanding mental task. The respondent must understand what is being asked, retrieve relevant information from memory, combine it into a judgement and then fit that judgement to the answer options offered (Tourangeau et al., 2000). Each step can fail. If the question is ambiguous, different respondents answer different questions. If it asks for information people do not have, they guess. If it signals a preferred answer, some will give it.

Because these failures are systematic, they produce bias rather than random noise. Everyone who misreads a term in the same way pushes the results in the same direction. Larger samples do not help; they only make the wrong answer more precise. That is why methodologists regard question design as one of the most important determinants of survey quality, and why the main reviews of the field treat it as a science with a substantial experimental base (Schaeffer & Presser, 2003; Schaeffer & Dykema, 2020).

Clarity

A clear question is one that every respondent understands in the same way, and in the way the researcher intended. That is harder to achieve than it sounds.

Words that seem clear but are not

Everyday words often carry several meanings. Fowler (1992) tested questions containing apparently simple terms, such as "exercise" or "servings of fat," and found that respondents interpreted them in markedly different ways. When the terms were defined or the questions rewritten more concretely, the answers changed substantially. William Belson's studies in Britain went further: after asking people to explain how they had understood a series of ordinary survey questions, he found that a large share of respondents had interpreted at least part of each question differently from the researcher's intention (Belson, 1981).

Words such as "regularly," "often," "household," "income," "family," "recently" and "use" are frequent culprits. Does "household income" include the earnings of a lodger? Does "recently" mean last week or last year? Does a respondent who "uses" a product once a year count as a user? Each of these questions should be answered in the question itself, not left to respondents.

Practical rules for clarity

The classic guides converge on a short list of rules (Payne, 1951; Converse & Presser, 1986; Bradburn et al., 2004):

- **Use simple, familiar words.** Prefer "help" to "facilitate" and "about" to "regarding." Jargon that is familiar inside an organisation is often unfamiliar outside it.
- **Be specific about the time period.** "In the past 30 days" is better than "recently."
- **Define terms that matter.** If the answer depends on what counts, say what counts.
- **Ask about one thing at a time.** A question such as "How satisfied are you with the price and quality of our service?" cannot be answered by someone who likes the quality but not the price. Such double-barrelled questions are among the most common faults in practice.
- **Avoid negatives, especially double negatives.** "Do you agree that the company should not stop offering refunds?" forces respondents to untangle the logic before they can answer.
- **Make the task clear.** Tell respondents what form the answer should take: a number, a choice from a list, a rating.

Research on question comprehension supports these rules. Graesser et al. (2006) developed a computer tool that flags features known to cause comprehension problems, such as unfamiliar technical terms, vague or imprecise relative terms, vague noun phrases, complex syntax and heavy demands on working memory. Lenzner et al. (2010) showed experimentally that questions containing such features take longer to answer, a sign of greater cognitive burden, and Holbrook et al. (2006) found that respondents had more trouble understanding and answering questions that were longer or used more complex wording.

Brevity

Brevity applies at two levels: the individual question and the questionnaire as a whole.

Short questions

Shorter questions are generally easier to understand and remember, especially when they are read aloud by an interviewer. A useful rule of thumb from the early guides is that most questions should be no more than about twenty words (Payne, 1951; Oppenheim, 1992). Brevity is not an absolute rule, however. Adding a short introduction or definition can improve understanding, and for questions about behaviour, longer questions that give memory cues can produce more complete reports than terse ones (Bradburn et al., 2004). The real principle is economy: every word should help the respondent understand and answer, and words that do not should go.

Short questionnaires

At the level of the whole questionnaire, length affects both who responds and how well they respond. Galesic and Bosnjak (2009) randomly told invited respondents that a web survey would take 10, 20 or 30 minutes. The longer the stated length, the fewer people started the survey. Among those who took part, questions placed later in the questionnaire were answered faster and with shorter open-ended answers and more uniform ratings, all signs of declining effort. Krosnick (1991) described this kind of shortcut-taking as satisficing: when the effort of answering carefully exceeds a respondent's motivation, they settle for an answer that is merely acceptable.

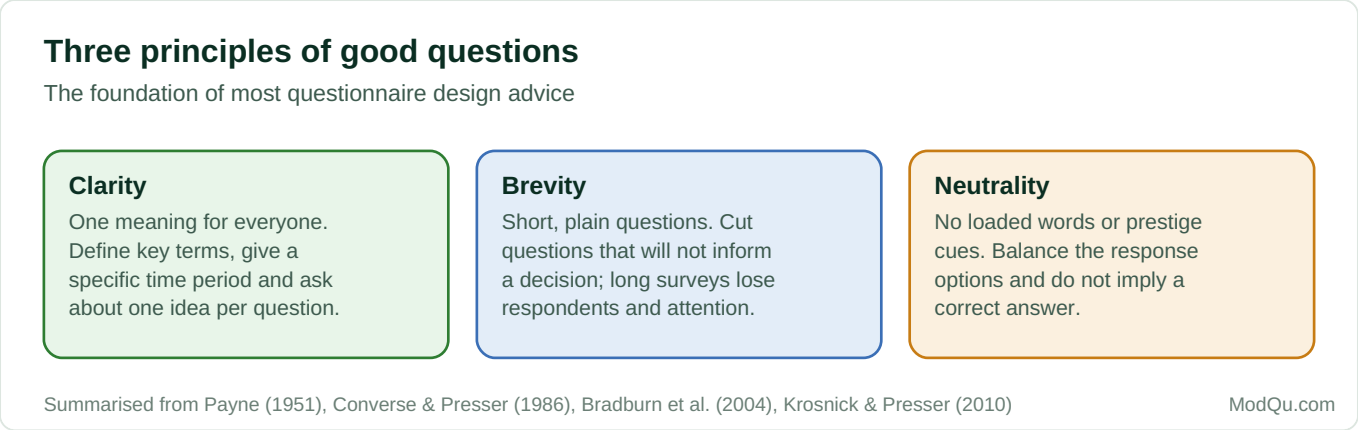


Figure 1. The three core principles of question writing. Summarised from Payne (1951), Converse and Presser (1986), Bradburn et al. (2004) and Krosnick and Presser (2010). Diagram by ModQu.

The practical lesson is to ask only what is needed. A useful discipline is to require a reason for every question: what decision or analysis will it inform? Questions that are "interesting to know" but have no planned use add burden and reduce the quality of the answers to the questions that matter.

Neutrality

A neutral question does not push respondents towards any particular answer. Departures from neutrality can be obvious, as in a question that begins "Don't you agree that...", or subtle, arising from a single word or an unbalanced set of options.

Loaded words

Some of the best-known findings in survey research concern single words. In one early experiment, many more Americans said the government should "not allow" public speeches against democracy than said it should "forbid" them, even though the two phrasings describe the same policy (Rugg, 1941). Later experiments showed far more public support for spending on "assistance to the poor" than for spending on "welfare" (Smith, 1987; Rasinski, 1989). The terms differ not in literal meaning but in the associations they carry, and those associations shape the answers.

Memory research shows a related effect. When people who had watched a film of a car accident were asked how fast the cars were going when they "smashed into" each other, they gave higher speed estimates than those asked how fast the cars were going when they "hit" each other, and were more likely to recall seeing broken glass that was not there (Loftus & Palmer, 1974). Words in a question can shape not only how people answer but what they remember.

Balanced options and formats

Neutrality also depends on the response options. A scale running from "good" to "excellent," with no negative options, pushes answers upwards. A question that names only one side of an issue ("Do you favour the new policy?") tends to attract more agreement than one that names both sides ("Do you favour or oppose the new policy?").

Agree-disagree formats carry a particular risk. Some respondents tend to agree with statements regardless of their content, a tendency known as **acquiescence**. Questions that ask directly about the attribute of interest, with response options specific to it, avoid this problem. Saris et al. (2010) compared agree-disagree questions with such item-specific questions across many studies and found that item-specific versions generally produced data of higher quality.

Prestige and social desirability

Questions can also become non-neutral by invoking authority or social norms. Mentioning that a policy is supported by experts, or that most people do something, signals an expected answer. On sensitive topics, people tend to give answers that present themselves favourably, so questions should be worded to make every answer acceptable and, where possible, asked in a self-administered mode (Tourangeau & Yan, 2007).

Beyond wording: options, order and mode

The three principles apply not only to the words of a question but to everything that surrounds it.

Response options

Closed questions are only as clear as their answer categories. Categories should be **exhaustive**, so that every respondent can find an answer that fits, and **mutually exclusive**, so that no respondent fits two. Age bands such as "18–25" and "25–35" break the second rule; a list of reasons for leaving a supplier without an "other" option breaks the first. When categories are numeric, their range itself sends a message. People asked how many hours of television they watched per

day reported more viewing when the scale's categories ran to high numbers than when they stopped at lower ones, apparently because they used the range as a clue to what was normal (Schwarz et al., 1985). Clarity and neutrality therefore apply to the options as much as to the stem of the question.

The number of scale points and their labels also matter. Fully labelled scales, in which every point carries words rather than only the end points, tend to be understood more consistently. For bipolar concepts such as satisfaction, five or seven points with a clear midpoint are a common choice, while unipolar concepts such as importance work better on scales that run from "not at all" to "extremely" (Krosnick & Presser, 2010).

Question order and context

Questions are not answered in isolation. An earlier question can change how a later one is interpreted or what comes to mind when it is answered. Schuman and Presser (1981) documented many such context effects; in one well-known case, support for allowing communist reporters into the United States rose sharply when the question followed one about allowing American reporters into the Soviet Union, because the first question invoked a norm of fairness. General questions, such as overall satisfaction, are usually best asked before specific ones, so that the answers to the specific items do not shape the general judgement. When order effects are likely and no order is obviously better, randomising the order across respondents spreads the effect evenly.

The mode of delivery

A question that works on a screen may not work when it is read aloud, and the reverse. In self-administered modes, respondents can see all the options and reread the question; long lists are manageable but early options tend to be chosen more often. In interviewer-administered modes, respondents must hold the options in memory, which favours short lists and simple scales, and later options tend to be chosen more often. Dillman et al. (2014) recommend writing questions that work in the same way across modes when a survey uses more than one, and keeping the visual layout of web and paper questionnaires simple and consistent, since layout itself communicates meaning to respondents.

The three principles together

Clarity, brevity and neutrality support one another but sometimes pull in different directions. Defining a term adds words. Presenting both sides of an issue makes a question longer. Removing a loaded word may require a more technical alternative that is less familiar. Krosnick and Presser (2010) emphasise that good design involves balancing these demands using evidence about how respondents actually process questions, rather than applying any single rule mechanically.

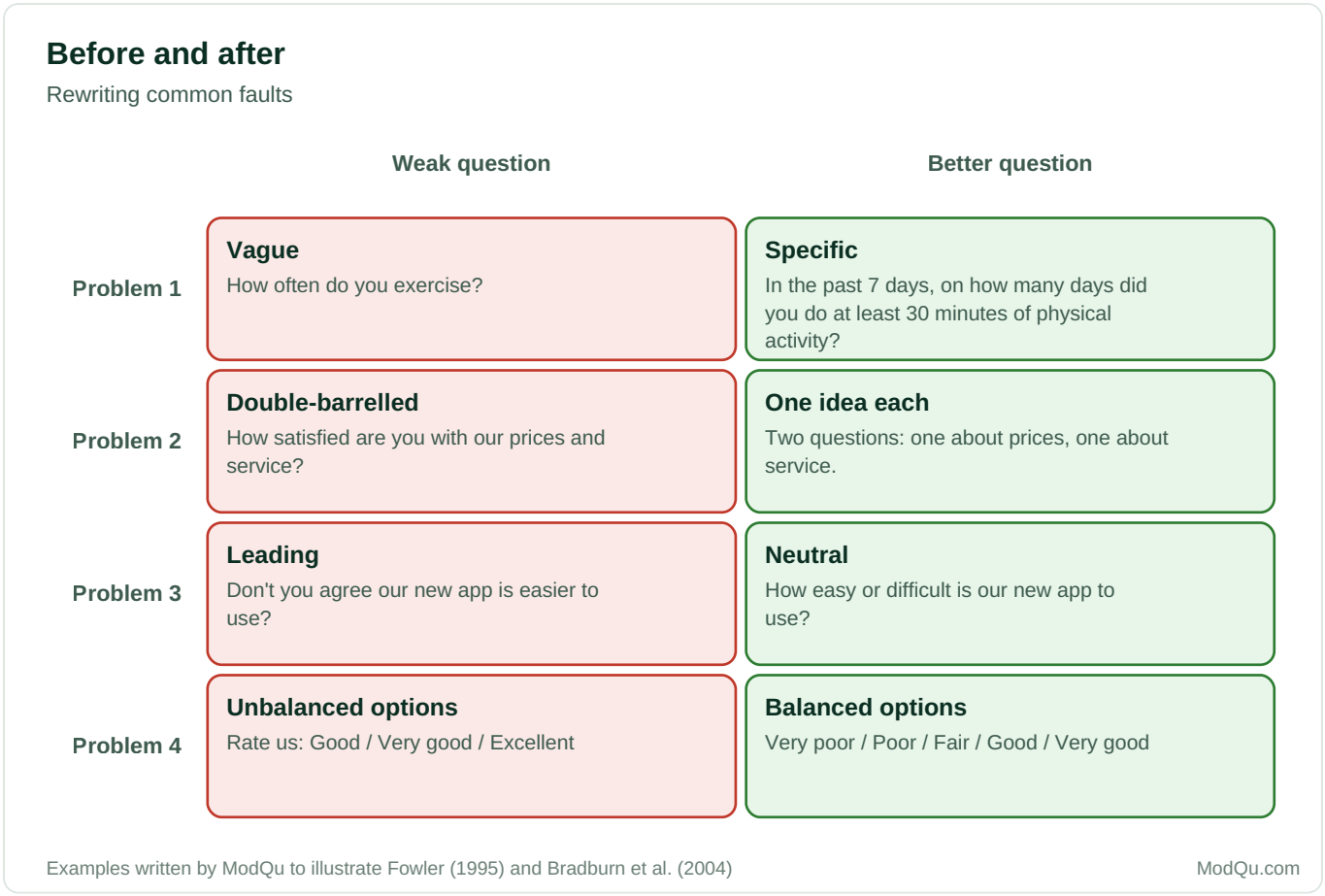


Figure 2. Common flaws and better alternatives. Examples written by ModQu to illustrate the principles in Fowler (1995) and Bradburn et al. (2004). Diagram by ModQu.

Testing, not guessing

Because question writers cannot reliably predict how others will read their questions, the principles must be backed by testing. Several methods are well established (Presser et al., 2004):

- **Expert review**, in which experienced questionnaire designers check each question against known problems.
- **Cognitive interviewing**, in which a small number of people from the target population answer the questions while thinking aloud or responding to follow-up probes about how they understood them.
- **Pilot testing**, in which the full questionnaire is administered to a small sample under realistic conditions to check flow, length and technical function.
- **Split-sample experiments**, in which different versions of a question are given to random halves of a sample to measure the effect of the wording directly.

Fowler (1995) argues that every important survey question should be evaluated before it is used, and that even simple pretesting with a handful of respondents identifies many of the most serious problems.

Applying the principles in business surveys

The same principles apply with particular force in customer and employee research, where questionnaires are often drafted by people close to the product or the organisation. Internal vocabulary is a common source of confusion: product names, team names and abbreviations that are familiar inside a company may mean nothing to customers. Enthusiasm for a new initiative can creep into wording ("How much has our improved service helped you?"). And the temptation to add "just one more question" for each interested department steadily lengthens questionnaires until respondents stop paying attention.

Tracking studies add a further consideration. Once a question has been used to measure a trend, improving its wording creates a break in the series, because the new version may produce different answers even if nothing has changed in the population. Where a flawed question must be fixed, the old and new versions should be asked side by side, with random halves of the sample, for at least one round so that the size of the wording effect can be measured. The same caution applies to changes in response options, scale labels and the position of a question in the questionnaire, each of which can shift results on its own (Schuman & Presser, 1981).

A short checklist helps. For each question, ask: Could anyone reasonably read this in two ways? Does it ask about more than one thing? Does it suggest a preferred answer? Are the response options balanced and complete? Will the answer be used? Questions that fail any of these tests should be rewritten or removed before the survey goes into the field.

Conclusion

The principles of clarity, brevity and neutrality are simple to state and surprisingly hard to follow. Clear questions mean the same thing to everyone; brief questions and questionnaires respect respondents' time and attention; neutral questions let respondents' own views come through. The research behind these principles shows that departures from them are not minor stylistic lapses but sources of systematic error that can change the conclusions of a study. Writing good questions is therefore not a preliminary to the real work of a survey. It is the real work, and it deserves the same care as the sample design and the analysis.

References

Belson, W. A. (1981). *The design and understanding of survey questions*. Gower.

Bradburn, N. M., Sudman, S., & Wansink, B. (2004). *Asking questions: The definitive guide to questionnaire design for market research, political polls, and social and health questionnaires* (Rev. ed.). Jossey-Bass.

Converse, J. M., & Presser, S. (1986). *Survey questions: Handcrafting the standardized questionnaire*. SAGE. https://doi.org/10.4135/9781412986045

Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method* (4th ed.). Wiley.

Fowler, F. J., Jr. (1992). How unclear terms affect survey data. *Public Opinion Quarterly, 56*(2), 218–231. https://doi.org/10.1086/269312

Fowler, F. J., Jr. (1995). *Improving survey questions: Design and evaluation*. SAGE.

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. https://doi.org/10.1093/poq/nfp031

Graesser, A. C., Cai, Z., Louwerse, M. M., & Daniel, F. (2006). Question Understanding Aid (QUAID): A web facility that tests question comprehensibility. *Public Opinion Quarterly, 70*(1), 3–22. https://doi.org/10.1093/poq/nfj012

Holbrook, A., Cho, Y. I., & Johnson, T. (2006). The impact of question and respondent characteristics on comprehension and mapping difficulties. *Public Opinion Quarterly, 70*(4), 565–595. https://doi.org/10.1093/poq/nfl027

Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 263–313). Emerald.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Lenzner, T., Kaczmirek, L., & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology, 24*(7), 1003–1020. https://doi.org/10.1002/acp.1602

Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13*(5), 585–589. https://doi.org/10.1016/S0022-5371(74)80011-3

Oppenheim, A. N. (1992). *Questionnaire design, interviewing and attitude measurement* (New ed.). Pinter.

Payne, S. L. (1951). *The art of asking questions*. Princeton University Press.

Presser, S., Couper, M. P., Lessler, J. T., Martin, E., Martin, J., Rothgeb, J. M., & Singer, E. (2004). Methods for testing and evaluating survey questions. *Public Opinion Quarterly, 68*(1), 109–130. https://doi.org/10.1093/poq/nfh008

Rasinski, K. A. (1989). The effect of question wording on public support for government spending. *Public Opinion Quarterly, 53*(3), 388–394. https://doi.org/10.1086/269158

Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly, 5*(1), 91–92. https://doi.org/10.1086/265467

Saris, W. E., Revilla, M., Krosnick, J. A., & Shaeffer, E. M. (2010). Comparing questions with agree/disagree response options to questions with item-specific response options. *Survey Research Methods, 4*(1), 61–79. https://doi.org/10.18148/srm/2010.v4i1.2682

Schaeffer, N. C., & Dykema, J. (2020). Advances in the science of asking questions. *Annual Review of Sociology, 46*, 37–60. https://doi.org/10.1146/annurev-soc-121919-054544

Schaeffer, N. C., & Presser, S. (2003). The science of asking questions. *Annual Review of Sociology, 29*, 65–88. https://doi.org/10.1146/annurev.soc.29.110702.110112

Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording, and context*. Academic Press.

Schwarz, N., Hippler, H.-J., Deutsch, B., & Strack, F. (1985). Response scales: Effects of category range on reported behavior and comparative judgments. *Public Opinion Quarterly, 49*(3), 388–395. https://doi.org/10.1086/268936

Smith, T. W. (1987). That which we call welfare by any other name would smell sweeter: An analysis of the impact of question wording on response patterns. *Public Opinion Quarterly, 51*(1), 75–83. https://doi.org/10.1086/269015

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge University Press. https://doi.org/10.1017/CBO9780511819322