RESEARCH · RESEARCH DESIGN

The Role of Pilot and Feasibility Studies in De-Risking Larger Research

What pilot and feasibility studies are for, how they reduce the risk that a large study will fail, what they should and should not try to estimate, how to set progression criteria, sample sizes for pilots, and how the same logic applies to surveys and commercial research.

ModQu research article · 2,751 words · 24 references (APA 7th edition)

Large studies are expensive, and many fail for reasons that have nothing to do with the research question. Participants cannot be recruited fast enough, questionnaires are misunderstood, an intervention cannot be delivered as planned, or data collection costs twice the budget. A **pilot** or **feasibility study** is a smaller study conducted beforehand to find these problems while they can still be fixed. Thabane et al. (2010) describe pilot studies as a way to assess the feasibility of a future study, reduce the risk of failure and improve its quality and efficiency. This article explains what such studies are for, what they can and cannot tell researchers, how to design and judge them, and how the same approach applies to surveys and commercial projects.

Why large studies fail

The reasons large studies fail are often practical. McDonald et al. (2006) reviewed 114 randomised trials funded by two United Kingdom agencies and found that fewer than a third recruited their original target number of participants within the planned time, and about half had to extend their recruitment period. Treweek et al. (2013) reviewed strategies to improve recruitment and found that few had strong evidence behind them, which makes testing recruitment in advance all the more important. Chalmers and Glasziou (2009) estimated that a large share of research investment is wasted through poor design, failure to complete and inadequate reporting; many of these failures could be anticipated by testing the study on a small scale first.

Pilot or feasibility?

The terms are often used interchangeably, and Arain et al. (2010) found considerable inconsistency in how researchers and journals used them. Eldridge, Lancaster et al. (2016) developed a conceptual framework, through a consensus process, that has become widely adopted.

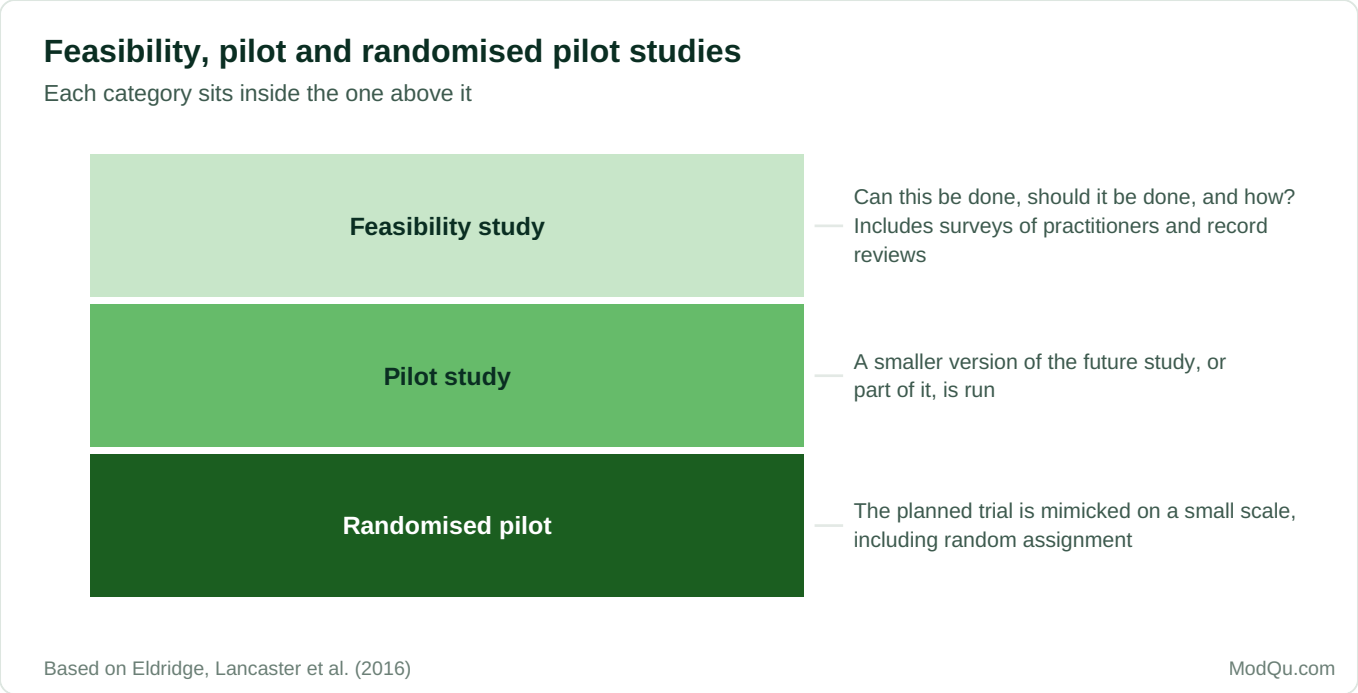


Figure 1. Feasibility, pilot and randomised pilot studies. Pilot studies are one type of feasibility study, and randomised pilots one type of pilot. Based on Eldridge, Lancaster et al. (2016); diagram by ModQu.

In their framework, a **feasibility study** asks whether something can be done, whether it should be done and, if so, how. It is the broad category. A **pilot study** is a type of feasibility study in which a future study, or part of it, is conducted on a smaller scale. A **randomised pilot study** mimics the planned randomised trial on a small scale, including random assignment. **Non-randomised pilot studies** test some parts of the planned study, such as recruitment or data collection, without randomisation. Other feasibility studies do not reproduce the future study at all; they might, for example, survey practitioners to see whether an intervention would be acceptable, or examine records to estimate how many eligible participants exist.

A pilot can be **external**, a separate study whose data are not used in the main analysis, or **internal**, the first phase of the main study, with predefined criteria for deciding whether to continue (Avery et al., 2017).

What a pilot should examine

Bowen et al. (2009) proposed several areas of focus for feasibility studies, including acceptability, demand, implementation, practicality, adaptation, integration, expansion and limited efficacy testing. In practice, most pilots examine some combination of the following.

Recruitment. How many eligible people are there, what proportion agree to take part, how long does it take to recruit them, and which recruitment methods work?

Retention and adherence. How many participants complete the study, how many stay in their assigned group and how many follow the intervention as intended?

Measures. Do participants understand the questions? Are the measures reliable, sensitive to change and acceptable? How much data is missing?

Procedures. Do randomisation, data collection, data entry and other procedures work as planned? How long do they take?

The intervention. Can it be delivered as intended by the people who will deliver it in the main study? Is it acceptable to participants?

Resources. How much does each part of the study cost in time, money and staff?

Lancaster et al. (2004) recommended that pilot studies have clear objectives, a protocol and a plan for analysis, just like any other study, and that their results be reported so that others can learn from them. Eldridge, Chan et al. (2016) published an extension of the CONSORT reporting guideline specifically for randomised pilot and feasibility trials.

What a pilot should not do

A common misuse of pilot studies is to test whether an intervention works. Leon et al. (2011) explained why this is a mistake. Pilot studies are small, so their estimates of an intervention's effect are very imprecise. A pilot that finds no significant effect does not show that the intervention is ineffective, and one that finds a large effect may simply have been lucky. Kistin and Silverstein (2015) warned that pilot results are often presented as preliminary evidence of effectiveness, which they cannot provide.

A related misuse is to estimate the effect size from a pilot and use it to calculate the sample size for the main study. Kraemer et al. (2006) showed that this practice is likely to lead to inappropriate decisions. Because pilot estimates are so imprecise, the calculated sample size can be wildly too large or too small, and studies of promising interventions may be abandoned because the pilot happened to underestimate the effect.

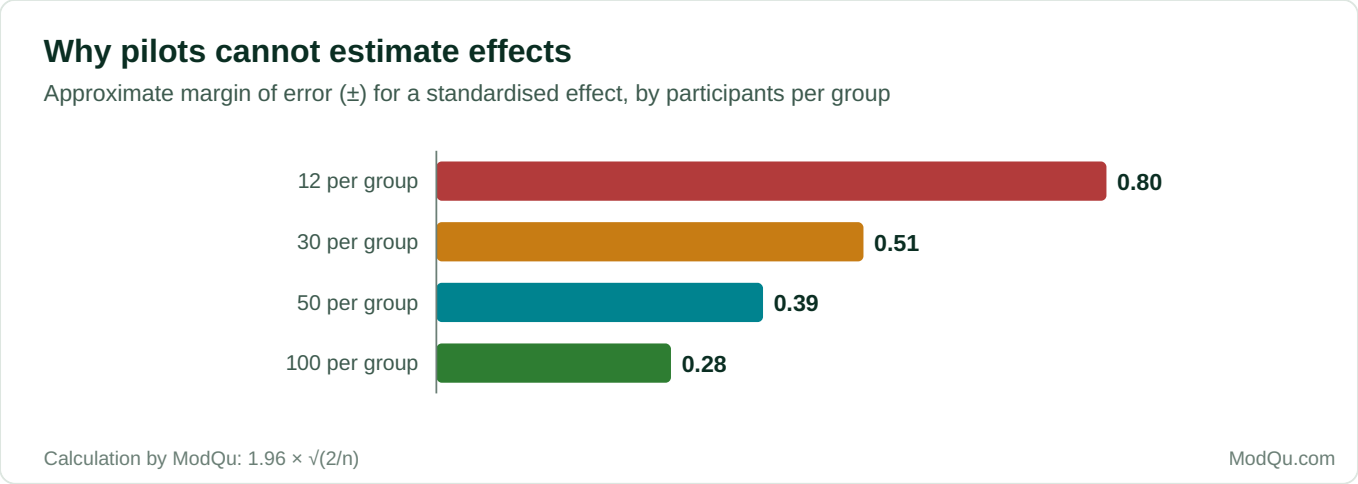


Figure 2. How imprecise pilot effect estimates are. Bars show the approximate half-width of a 95% confidence interval for a standardised mean difference, by number of participants per group (normal approximation). Calculation by ModQu.

The figure shows why. With 12 participants per group, a common pilot size, the 95% confidence interval for a standardised effect extends roughly 0.8 standard deviations either side of the estimate. An observed effect of 0.3 is consistent with anything from a moderate harm to a large benefit. Even with 50 per group, the interval spans almost 0.8 standard deviations. Leon et al. (2011) recommended that the sample size for the main study be based on the smallest effect that would be clinically or practically meaningful, not on a pilot estimate.

Pilots are, however, useful for estimating quantities other than the effect itself, such as the standard deviation of the outcome, the recruitment rate, the dropout rate and, in cluster studies, the degree of similarity among participants within clusters. These **nuisance parameters** feed directly into sample size calculations and budgets. Browne (1995) showed that variances estimated from small pilots are also imprecise and recommended using an upper confidence limit, rather than the point estimate, when planning.

How large should a pilot be?

There is no single answer, because the size should depend on what the pilot is meant to estimate. Several rules of thumb have been proposed. Julious (2005) suggested 12 participants per group as a minimum for estimating a mean and variance with reasonable precision. Hertzog (2008) discussed the trade-offs and suggested that 10 to 40 per group may be appropriate, depending on the aims. Sim and Lewis (2012) argued that pilot size should be calculated in relation to the precision needed and the efficiency of the overall programme, and suggested that larger pilots, often 50 or more per group, may be justified to estimate the standard deviation reliably. Whitehead et al. (2016) showed how to choose the pilot size so as to minimise the total number of participants across the pilot and main trial, finding that the optimal pilot size depends on the effect size the main trial aims to detect.

For feasibility objectives such as recruitment and retention rates, the required size depends on how precisely the rate must be known. Estimating a retention rate of around 80% to within plus or minus 10 percentage points requires roughly 60 participants.

Progression criteria

A pilot is only useful if its results lead to a decision. **Progression criteria** set out in advance what results would lead the team to proceed with the main study as planned, proceed with changes, or stop. Avery et al. (2017) found that many pilot studies did not specify such criteria, and recommended that they be agreed with funders before the pilot begins.

A common approach uses a traffic light system for each key feasibility measure. For recruitment, for example:

- **Green:** at least 80% of the target recruitment rate is achieved; proceed.
- **Amber:** 50% to 80% is achieved; proceed with changes to recruitment methods.
- **Red:** less than 50% is achieved; do not proceed without major redesign.

Setting criteria in advance prevents the temptation to reinterpret disappointing pilot results as acceptable, and makes clear to funders and stakeholders what the pilot is meant to establish.

An illustrative example

Consider a hypothetical pilot for a twelve-month study of a workplace wellbeing programme, run in two offices before a planned roll-out to twenty. The numbers below are illustrative. The team agreed four progression criteria in advance: recruitment, retention at three months, completion of the outcome questionnaire, and delivery of sessions as designed.

- **Recruitment:** the target was 60 staff in eight weeks; 41 enrolled, 68% of target. This is amber, so the team proceeds with changes.
- **Retention at three months:** 35 of 41 remained, 85%. This is green.
- **Questionnaire completion:** 33 of 35 completed every item, 94%, but cognitive interviews showed that "workload" was read in two different ways. The team proceeds with one item reworded.
- **Delivery:** facilitators delivered 9 of 12 planned sessions in full in one office and 5 of 12 in the other, because rooms were not available. This is red for the second site, and the team must book rooms before sessions are scheduled.

None of these results says whether the programme improves wellbeing. They do say that the main study, as first designed, would have under-recruited, used an ambiguous measure and suffered uneven delivery. Each problem has a specific, inexpensive fix. That is the purpose of a pilot.

Internal and external pilots

An **external pilot** is a separate study whose data are not used in the main analysis. An **internal pilot** is the first phase of the main study itself: data collected during the pilot phase are carried forward into the final analysis, and the study continues if the progression criteria are met (Eldridge, Lancaster et al., 2016). Internal pilots save time and participants, because nothing is discarded. They also have a drawback: the design cannot change much after the pilot phase without compromising the main analysis, so they suit studies where the main uncertainty is recruitment rather than the intervention or the measures.

External pilots allow larger changes. If the pilot shows that a questionnaire must be rewritten or the intervention redesigned, the main study can start again with the revised version. In practice, teams developing new interventions or new measures tend to need an external pilot, while teams running a well-understood intervention in a new population may need only an internal one. Kraemer et al. (2006) add a related caution: when the decision to proceed depends on preliminary effect estimates, a promising study may be abandoned, or a weak one funded, on the basis of numbers too imprecise to support either decision.

Pilots for complex interventions

Many interventions in health, education and social policy involve several interacting components, multiple groups of people and considerable flexibility in delivery. The Medical Research Council guidance on complex interventions treats feasibility and piloting as a distinct phase in the development and evaluation of such interventions (Craig et al., 2008). At this stage, researchers test whether the intervention can be delivered, whether it is acceptable, how it should be adapted

for different settings and how outcomes should be measured, before moving to a full evaluation. Moore et al. (2011) recommended that pilot studies in clinical and translational research focus explicitly on such questions of process and feasibility rather than on preliminary effects.

Pretesting surveys

In survey research, the equivalent of a pilot is the **pretest**. Presser et al. (2004) reviewed methods for testing survey questions, including expert review, **cognitive interviewing**, behaviour coding of interviews, and field pretests with small samples. Willis (2005) describes cognitive interviewing in detail: respondents answer draft questions while thinking aloud or responding to probes about how they understood each question, how they recalled the information and how they chose their answer. This reveals problems that are invisible in the answers themselves, such as respondents interpreting "household" differently or being unable to recall the information requested.

A field pretest then runs the full questionnaire with a small sample under real conditions, checking the length of the interview, the flow of questions, skip patterns, the distribution of answers and the rate of missing data. van Teijlingen and Hundley (2001) note that pilot studies can reveal problems in questionnaires, procedures and logistics that would be costly or impossible to correct once the main study has begun.

Pilots in commercial and organisational research

The logic of piloting applies directly to commercial research and to organisational change. Before launching a national customer survey, a firm can pretest the questionnaire with a few dozen customers to find confusing questions, measure completion time and check that the sample source delivers the expected response rate. Before rolling out a new service model across all branches, an organisation can pilot it in a few to check whether staff can deliver it, whether customers accept it and what it costs.

In product development, the same idea underlies the **minimum viable product**, which Ries (2011) describes as the version of a new product that allows a team to collect the maximum amount of validated learning about customers with the least effort. Like a feasibility study, it is designed to answer questions about whether something can work, and whether to proceed, before committing large resources.

The same cautions apply. A small commercial pilot is excellent for finding practical problems but poor at estimating the size of an effect. If a pilot of a new loyalty scheme in three stores shows a 5% increase in sales, the result is far too uncertain to forecast national revenue. The pilot's real value lies in showing whether the scheme can be run as intended, how customers respond to it and what it costs, and in setting clear criteria for a larger test.

Reporting a pilot

Pilot studies are often unpublished or reported briefly, which means that other teams repeat the same mistakes. Chalmers and Glasziou (2009) estimated that a large share of research investment is wasted through poor design and incomplete reporting, and unreported pilots contribute to that waste. The CONSORT extension for pilot and feasibility trials sets out what a report should contain (Eldridge, Chan et al., 2016). Its main changes to the standard checklist are:

- state the feasibility objectives, not hypotheses about effectiveness;
- give the rationale for the pilot's sample size;
- report the progression criteria and whether each was met;
- present any outcome estimates descriptively, with confidence intervals, and avoid hypothesis tests of effect;
- describe what changes, if any, will be made to the main study.

A report of this kind is short, but it records the decisions and the evidence for them. It also allows funders to see that the main study rests on tested procedures.

Common mistakes

Reviews of published pilot studies find recurring problems (Arain et al., 2010; Kistin & Silverstein, 2015). The most common are:

1. **No stated feasibility objectives.** The pilot is described as "preliminary" without saying what it is meant to establish.
2. **Testing effectiveness.** The report leads with a p value for the main outcome from a sample far too small to estimate it.
3. **No progression criteria.** Results are interpreted after the fact, and almost any outcome is judged acceptable.
4. **Unrepresentative settings.** The pilot runs at the most enthusiastic site with the most experienced staff, so recruitment and delivery look easier than they will be elsewhere.
5. **No time to act.** The main study is scheduled to start before the pilot results are analysed, so lessons cannot be applied.

Each of these can be avoided at the planning stage, at little cost.

Conclusion

Pilot and feasibility studies reduce the risk of large studies by testing whether they can be done before committing to doing them. They are most valuable when they focus on feasibility: recruitment, retention, measures, procedures, delivery and cost. They are least valuable, and potentially misleading, when used to test whether an intervention works or to estimate its effect for sample size calculations. Clear objectives, a protocol, appropriate sample sizes and progression criteria agreed in advance turn a pilot from an informal trial run into a disciplined step that makes the main study more likely to succeed.

References

Arain, M., Campbell, M. J., Cooper, C. L., & Lancaster, G. A. (2010). What is a pilot or feasibility study? A review of current practice and editorial policy. *BMC Medical Research Methodology, 10*, Article 67. https://doi.org/10.1186/1471-2288-10-67

Avery, K. N. L., Williamson, P. R., Gamble, C., O'Connell Francischetto, E., Metcalfe, C., Davidson, P., Williams, H., & Blazeby, J. M. (2017). Informing efficient randomised controlled trials: Exploration of challenges in developing progression criteria for internal pilot studies. *BMJ Open*, *7*(2), Article e013537. https://doi.org/10.1136/bmjopen-2016-013537

Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., Weiner, D., Bakken, S., Kaplan, C. P., Squiers, L., Fabrizio, C., & Fernandez, M. (2009). How we design feasibility studies. *American Journal of Preventive Medicine*, *36*(5), 452–457. https://doi.org/10.1016/j.amepre.2009.02.002

Browne, R. H. (1995). On the use of a pilot sample for sample size determination. *Statistics in Medicine*, *14*(17), 1933–1940. https://doi.org/10.1002/sim.4780141709

Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *The Lancet*, *374*(9683), 86–89. https://doi.org/10.1016/S0140-6736(09)60329-9

Craig, P., Dieppe, P., Macintyre, S., Michie, S., Nazareth, I., & Petticrew, M. (2008). Developing and evaluating complex interventions: The new Medical Research Council guidance. *BMJ*, *337*, Article a1655. https://doi.org/10.1136/bmj.a1655

Eldridge, S. M., Chan, C. L., Campbell, M. J., Bond, C. M., Hopewell, S., Thabane, L., & Lancaster, G. A. (2016). CONSORT 2010 statement: Extension to randomised pilot and feasibility trials. *BMJ*, *355*, Article i5239. https://doi.org/10.1136/bmj.i5239

Eldridge, S. M., Lancaster, G. A., Campbell, M. J., Thabane, L., Hopewell, S., Coleman, C. L., & Bond, C. M. (2016). Defining feasibility and pilot studies in preparation for randomised controlled trials: Development of a conceptual framework. *PLOS ONE, 11*(3), Article e0150205. https://doi.org/10.1371/journal.pone.0150205

Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing & Health*, *31*(2), 180–191. https://doi.org/10.1002/nur.20247

Julious, S. A. (2005). Sample size of 12 per group rule of thumb for a pilot study. *Pharmaceutical Statistics*, *4*(4), 287–291. https://doi.org/10.1002/pst.185

Kistin, C., & Silverstein, M. (2015). Pilot studies: A critical but potentially misused component of interventional research. *JAMA, 314*(15), 1561–1562. https://doi.org/10.1001/jama.2015.10962

Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavage, J. A. (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of General Psychiatry, 63*(5), 484–489. https://doi.org/10.1001/archpsyc.63.5.484

Lancaster, G. A., Dodd, S., & Williamson, P. R. (2004). Design and analysis of pilot studies: Recommendations for good practice. *Journal of Evaluation in Clinical Practice, 10*(2), 307–312. https://doi.org/10.1111/j.2002.384.doc.x

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research, 45*(5), 626–629. https://doi.org/10.1016/j.jpsychires.2010.10.008

McDonald, A. M., Knight, R. C., Campbell, M. K., Entwistle, V. A., Grant, A. M., Cook, J. A., Elbourne, D. R., Francis, D., Garcia, J., Roberts, I., & Snowdon, C. (2006). What influences recruitment to randomised controlled trials? A review of trials funded by two UK funding agencies. *Trials, 7*, Article 9. https://doi.org/10.1186/1745-6215-7-9

Moore, C. G., Carter, R. E., Nietert, P. J., & Stewart, P. W. (2011). Recommendations for planning pilot studies in clinical and translational research. *Clinical and Translational Science, 4*(5), 332–337. https://doi.org/10.1111/j.1752-8062.2011.00347.x

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

Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. Crown Business.

Sim, J., & Lewis, M. (2012). The size of a pilot study for a clinical trial should be calculated in relation to considerations of precision and efficiency. *Journal of Clinical Epidemiology, 65*(3), 301–308. https://doi.org/10.1016/j.jclinepi.2011.07.011

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L., & Goldsmith, C. H. (2010). A tutorial on pilot studies: The what, why and how. *BMC Medical Research Methodology, 10*, Article 1. https://doi.org/10.1186/1471-2288-10-1

Treweek, S., Lockhart, P., Pitkethly, M., Cook, J. A., Kjeldstrøm, M., Johansen, M., Taskila, T. K., Sullivan, F. M., Wilson, S., Jackson, C., Jones, R., & Mitchell, E. D. (2013). Methods to improve recruitment to randomised controlled trials: Cochrane systematic review and meta-analysis. *BMJ Open, 3*(2), Article e002360. https://doi.org/10.1136/bmjopen-2012-002360

van Teijlingen, E. R., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update, 35*. University of Surrey.

Whitehead, A. L., Julious, S. A., Cooper, C. L., & Campbell, M. J. (2016). Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Statistical Methods in Medical Research, 25*(3), 1057–1073. https://doi.org/10.1177/0962280215588241

Willis, G. B. (2005). *Cognitive interviewing: A tool for improving questionnaire design*. SAGE.