# PRACTICAL CONSIDERATIONS IN SAMPLE SIZE SELECTION FOR BEHAVIOR CODING PRETESTS

Andrew L. Zukerberg, Dawn R. Von Thurn and Jeffrey C. Moore, U.S. Bureau of the Census[1]
Andrew L. Zukerberg, SRD/CSMR, Room 3133 FOB 4, Washington, DC 20233-9150

## 1. BACKGROUND

Questionnaire designers have available to them a variety of pretest techniques, including debriefing, expert review, behavior coding and cognitive testing. Although it is nearly universal for survey textbooks and questionnaire designers to advocate pretesting questionnaires, few straightforward guidelines exist to determine a sample size which will maximize the efficiency of a pretest. Experts' recommendations about the appropriate size of questionnaire pretests cover a substantial range. For example, according to Sudman (1983), "A pilot test of 20 - 50 cases is usually sufficient to discover the major flaws in a questionnaire before they damage the main study" (p. 181). Sheatsley (1983) suggests a lower number: "It usually takes no more than 12 - 25 cases to reveal the major difficulties and weaknesses in a test questionnaire" (p. 226). However, Courtenay (1978) asserts that "for most purposes a pilot survey of between thirty and a hundred interviews is adequate. But the exact size will depend on the aims of the particular test: two or three interviewers doing five to ten interviews each will often be able to reveal wording and layout problems" (p. 51). Weinberg (1983), on the other hand, suggests that fairly large samples are needed for pretests: "Pretest conducted on reasonably large (100 or more) properly selected samples may be useful in deciding which items to keep and which to discard" (p. 332).

These suggestions obviously reflect a wealth of practical experience, but they are accompanied by virtually no empirical support. What holds for pretesting in general also holds for behavior coding as a specific technique of questionnaire pretesting. Very little research has been done to support the conventional wisdom about the number of cases needed for behavior coding pretests. Cannell, et. al. (1989), advocate what appears to be the approximate midpoint of the above recommendations: "We propose that the pretest be based on around 50 interviews with respondents similar to those to be selected in the final sample" (pg 84). However, they later provide a caveat to this recommendation: "While we recommend taking approximately 50 pretest interviews, that number may not be necessary to obtain adequate estimates of problems for questions with which the investigator has previous experience" (p. 89). DeMaio, et. al. (1993), recommend a larger number of cases for successful behavior coding: "Thus sample sizes for behavior coding may range from fifty cases to as many as several hundred" (pg. 26).

In this paper we describe a modest research project which offers some objective evidence about whether a quick, small-scale behavior coding pretest may serve a questionnaire designer's needs about as well as a substantially larger effort. There are, of course, extremely practical reasons for concern about the number of behavior coding cases necessary to be informative during a pretest. Each additional case means additional expense in the form of interviewer time, coder time, and material expenses. For the test we conducted, using Census Bureau cost figures, and factoring in field and professional staff time and overheads the cost of each additional 40 minute interview is roughly $150. A full accounting of behavior coding costs must also factor in schedule and deadline considerations, since each additional case interviewed and behavior coded adds time to a pretest. Again, using our test figures, 15 additional interviews could easily add 50 or more hours of work time to a pretest, and much more to the total duration of the test, depending on the availability of interviewers, behavior coders and respondents.

The wide range in recommended pretest sizes and the dearth of objective guidelines can be confusing for a researcher planning a pretest. Budget and time limitations may preclude large scale pretesting, but this doesn't offer much succor to the researcher concerned about doing quality work. We believe that a great deal can be learned about a questionnaire from a small number of behavior coded cases.

First, there is some objective evidence that a small-scale behavior coding study may yield a payoff commensurate with a larger effort. Presser and Blair (1994) analyzed half of the data from two behavior coding pretests of twenty-one and thirty cases. They found that the average number of problems identified in the halved datasets was roughly the same as in the full set of cases. This is, however, the only controlled study of this type which we could find.

The second reason for advocating behavior coding pretests that are more limited in scale is more philosophical. Behavior coding is a limited, qualitative tool for assessing likely questionnaire problems that are assumed to have implications for survey data quality. It shouldn't be the only technique utilized to revise or improve a survey questionnaire. Devoting too many resources and too much attention to qualitative studies in general may divert attention and resources away

from a more thorough and rigorous assessment of survey response errors using statistical samples and more direct and precise error detection techniques.

## 2. METHODOLOGY

### 2.1 Behavior Coding

Behavior coding is a technique which applies a frame of codes to the behaviors of interviewers and respondents while the interview occurs. Coding can be done concurrently with the interview or from a tape recording of the interview. Behavior coding has its roots in small group interaction coding. The technique was initially adapted to survey research as a tool for evaluating the effect of interviewer and respondent behaviors on reporting (Cannell et al., 1968). Recently, behavior coding has been used by researchers (e.g., Cannell and Robinson, 1971; Morton-Williams, 1979; Burgess and Paton, 1993) to identify questionnaire flaws through the coding of respondent-interviewer interactions. Behavior coding is less subjective than techniques such as interviewer debriefings or cognitive interviews because it applies codes to the interaction. The appeal of behavior coding is augmented by its flexibility (See Esposito et al., 1994). The complexity of coding schemes and level of interaction coded can be varied according to a researcher's goals.

### 2.2 The Research Context

This research into sample size requirements for behavior coding grew out of efforts to revise the American Housing Survey (AHS). The American Housing Survey is conducted by the Census Bureau for the Department of Housing and Urban Development. AHS provides information on the size and composition of the nation's housing inventory. For over a year, Census Bureau staff have been revising, testing, and re-revising subsections of the AHS instrument based on conclusions drawn from a variety of pretesting techniques, including behavior coding, cognitive interviews, interviewer debriefings, and expert reviews. The primary data for this paper come from tape recorded interviews using the current AHS form. (Our use of these recordings is somewhat tangential to their immediate purpose, which was to assess interviewer behaviors in their attempts to administer what we knew from cognitive interviews to be a particularly "unfriendly" section of the questionnaire for respondents.)

### 2.3 Methods

Twenty-five professional Census Bureau interviewers from five regional offices across the nation conducted the AHS interviews that we used for this test between November and December 1994. We selected a sample of 100 addresses from households which had participated in the 1990 wave of the survey and were selected for the 1994 enumeration. We were interested in English speaking, regular occupied households. Of the 100 households selected, 66 households met these requirements. Interviewers were asked to tape record the interview with the consent of the respondent. The primary focus of our work on AHS at the time was a series of questions about heating equipment; therefore, we selected cases in order to capture a range of heating equipment types, based on the type of heating equipment the household had reported in 1990. The six heating equipment types selected for inclusion were ones which indicated the greatest difficulty with the questions in a cognitive pretest. Given the age of the sample, we anticipated roughly 50 complete interviews; however, the 66 selected addresses produced only 31 usable taped interviews from 16 interviewers[2]

Our behavior coding focused on an initial section of the AHS questionnaire consisting of 38 questions. (Many of these questions were of minor substantive interest, but were included in the behavior coding for simplicity, to avoid requiring coders to have to search for the items of real interest.) The primary questions of interest in this section of the questionnaire dealt with a variety of subjects relating to the respondent's home and neighborhood, including questions on heating and cooling equipment; plumbing; number of rooms; physical problems such as peeling paint, and cracks and holes in floors, walls, and ceilings; and subjective ratings of the home and neighborhood. Question types included yes/no questions, ten point rating scales, and open-ended questions with respondent hand cards.

We coded the 31 taped interviews for first exchange level interaction only, using a scheme adapted from Cannell et al.(1989)[3]. First level of exchange consists of the interviewer reading the question and the respondent's initial response. We coded both interviewer and respondent behaviors. If an interviewer did not read a question exactly as worded or a respondent provided a non-adequate answer, notes were taken on the interaction. These notes provide substantive information on the questions. Two researchers behavior coded each interview independently; results were compared and all discrepancies were resolved by consensus.

Following our analysis of the AHS data, we replicated our analyses on a completely different data set derived from a computer assisted personal interview (CAPI) administered test of a decennial census related coverage measurement questionnaire. These interviews were conducted with 49 respondents, in a single urban site, by five professional Census Bureau interviewers. Each interview tape was behavior coded by one of three

behavior coders. The authors report inter-coder reliability levels of 81% (based on percent agreement) for interviewer codes and 79% for respondent codes, but there was no multiple coding and no attempt to reconcile coder differences (Bates and Kindred-Town, 1995).

## 3. FINDINGS

Our analysis of the behavior coding results focuses primarily on the simple question: What did we learn about questionnaire problems using the full set of available tape-recorded and behavior coded interviews that we would not have learned had we only gathered half (or even less than half) as many cases?

### 3.1 AHS Results

Our analysis of the primary AHS dataset omits 16 questions which, because of questionnaire skip patterns, yielded fewer than 10 question-and-response interactions (This is an obvious limitation of a small scale test, which we discuss later). We compared distributions of interviewer and respondent behavior codes for the remaining 22 questions on a question-by-question basis for the first 15 interviews conducted versus the full 31 cases, to see whether we gained information about questionnaire problems with the additional interviews[4].

Table 1 shows the interviewer behavior coding results for the first 15 cases and the full set of 31. To allow a more objective comparison between the reduced and full set of cases, we highlight in Table 1 all instances in which interviewers' "exact" and "slightly changed" readings of the question failed to exceed 85%. (This subjective 15% cutoff for problem identification was first suggested by Morton-Williams (1979). This rule has been adapted by others. Fowler (1989) states: "Although the selection of 15% was arbitrary, it proved a reasonably easy task to identify ambiguities and problems with questions that met those criteria" (p. 73). It now seems to be a fairly well-accepted standard among survey organizations analyzing behavior coding data.)

The "15% rule" identifies 9 items in the full 31-case data set as having caused interviewers problems. These same 9 items -- and 2 others -- are also identified as problems with only 15 cases. The overall picture is one of striking similarity. The only knowledge "gained" from the second half of our behavior coding exercise was that two items which would perhaps have been labelled "problems" in the initial set of cases (one is right on the cutoff) were not so labelled in the full set of cases.

Table 2 presents the results for the respondent behavior codes. Again, we applied the "15% rule" to identify items which we would label as having caused problems for respondents. For both the first 15 cases and the entire set of 31 the same eleven questions are identified as problematic; the additional 16 behavior coded cases added no new knowledge about items that caused problems for respondents. Interestingly, despite the fact that both the interviewer and respondent results implicated about half the 22 items as problematic, there were only 4 items common to both "problem" subsets.

Using a strict cutoff is often a first step in identifying problematic questions when looking at behavior coding results. In practice, however, researchers almost always want to delve further into the data to try to understand the nature of the problems that the behavior coding has flagged. We looked at the types of behaviors coded for each question to see if we would have ascertained different information from the full number of cases than with the halved data set. We found that the behaviors displayed by interviewers were virtually no different with the smaller or larger data set. Respondents' behaviors were only slightly less consistent. For two questions the total dataset identified more respondent behavior codes than did the halved data set; however, a closer examination of these differences suggests that the additional codes indicate similar problems with respondents' understanding of the intent of the question or the nature of the response task. For example, on one of the ratings questions, we identified both "qualified answers" and "interruptions" at both n=15 and n=31. At n=31 we also coded "request for clarification". While this may be new information, it may also indicate the same type of problem as is indicated by qualified answers, that is, uncertainty about what the question is asking.

In order to corroborate our analyses we asked four researchers who were not working on this project to serve as independent judges -- to review the behavior coding results and identify problematic questions. We did not suggest any specific decision rules. Two of the judges were given the data for the full 31 cases and the other two were given the results for only the initial fifteen cases. The two teams of judges responded very similarly. For 16 of the 22 items, whether or not the item was judged to be a problem was exactly the same for the team looking at only 15 cases and the team looking at all 31. (All four judges agreed on 12 of the 22 items; for another 4 items both teams gave a split decision, with one member of the team declaring the item to be a problem and the other not.) There were no completely opposing decisions -- that is, all of the remaining six items were labelled "problem" items by 3 of the 4 judges. When asked to select the five questions most in need of attention, the judges were unanimous on four of their selections. Not surprisingly,

these are the four questions for which behavior coding identified both respondent and interviewer problems at the 15% or greater level.

That similar conclusions are drawn by researchers working with 15 and 31 cases, confirms our earlier finding. Judges using different decision rules identified very similar sets of problematic questions regardless of whether their conclusions were drawn from the reduced or the full set of cases. The discrepancies which do exist are not due to the number of interviews conducted but to modest variations in the decision rules used by researchers.

### 3.2 Replication

We repeated our analyses on a completely different data set -- a CAPI-administered coverage measurement questionnaire that was being field tested for possible use in the decennial census. The interviews were conducted by five interviewers in a single urban site, with a total of 49 respondents. This dataset offered 10 questions for analysis. The researchers using this data set adopted a more liberal 20% cutoff for problem identification (Bates and Kindred-Town, 1995). We employed the same cutoff in our analysis of the data. Similar results were found. Behavior coding identified the same four questions as problematic for interviewers at 49 and 15 cases. And, the same five questions were problematic for respondents at n=15 and n=49.

## 4. DISCUSSION

A survey designer planning to use behavior coding to identify questionnaire items that need to be improved eventually must face the question, "How many cases do I need to behavior code?" A variety of factors will affect the answer to this question. Practical considerations, especially time and money resources, often are limiting factors. But in the absence of such constraints, there are only the most general guidelines to follow, and these guidelines generally suggest the need for a fairly substantial undertaking of perhaps 50 cases -- or even more.

Our practical needs first drove us to the literature in search of concrete guidance on behavior coding sample sizes. Since then, our own (admittedly limited) practical experience has made us wonder whether the general guidelines we found aren't overly conservative, and has led us to the belief that a great deal can be learned from a more modest-sized effort. That the knowledge gained from "large" (n=30 to 50) studies often differs only trivially from "small" (n=15) ones was born out in the present research on a dataset involving American Housing Survey interviews, and in a replication using a decennial census coverage measurement survey.

The results of this research can perhaps serve as a useful guide to others planning similar pretest efforts. Pretests can involve substantial time and money costs. We estimate, using Census Bureau figures for our test, a savings of about $5,000 and over 120 hours in time by conducting and behavior coding 15 interviews rather than 50. The time and money saved on a smaller pretest can allow a research program to test multiple iterations of a questionnaire rather than only one version.

We selected our sample of respondents by focusing on those with more problematic circumstances. By isolating and studying these respondents, we raised the likelihood of rapidly identifying questionnaire flaws.

Certainly there are circumstances which would demand a larger-scale effort than what we advocate here for general use, for example: (1) studies of questionnaires with skip patterns or other characteristics which result in questions of interest being administered in only a small subset of cases; (2) studies whose goals go well beyond the mere identification of the presence of problems, to an effort to fully understand the precise nature of those problems; (3) studies attempting to exploit the quantitative nature of the behavior coding technique by using it to detect the significant reduction of problems, through the comparison of coding results before and after questionnaire revisions; and (4) studies of questionnaires which produce established time series estimates, where question changes must be considered cautiously out of consideration for the continuity of the series. In most ordinary circumstances, however, we submit that a small behavior coding pretest can be as efficient at identifying questionnaire flaws as a larger one.

Clearly, there are limitations to this research. The high rate of failure to tape interviews erodes the quality of our sample. To address this we have requested that 50 additional interviews be taped during the 1995 administration of AHS. We will behavior code these and add them to our database.

We want to leave the right impression here. We do not advocate slipshod, quick-and-dirty research to identify and solve questionnaire problems; quite the opposite. We recommend small-scale behavior coding studies because we advocate more test-and-revise iterations, and more use of multiple qualitative methods to "triangulate" on solutions to questionnaire problems. Expending fewer scarce resources on one-shot behavior coding research will leave more resources for multiple iterations, multiple pretest techniques, and for the precise, quantitative research necessary for a truly rigorous assessment of survey response errors and the proposed solutions to them.

## References


Bates, N. and Kindred-Town, M. (1995). "The November Integrated Coverage Measurement (ICM) Test: Results from Behavior Coding of ICM Person Interviews." Internal Census Memorandum.

Burgess, M.J., and Paton, D. (1993). "Coding of Respondent Behavior by Interviewers to Test questionnaire Wording," in Proceedings of the Section on Survey Methods Research, Alexandria, VA: American Statistical Association, pp 392-397.

Cannell, C.F., Fowler, F.J., and Marquis, K.H. (1968) The Influence of Interviewer and Respondent Psychological and Behavioral Variables on the Reporting in Household Interviews. Washington: U.S. Government Printing Office.

Cannell, C., Oksenberg, L., Kalton, G., Bischoping, K. and Fowler, F.J. (1989). "New techniques for Pretesting Survey Questions." Research Report. Survey Research Center, The University of Michigan.

Cannell, C.F. and Robinson, S. (1971). "Analysis of Individual Questions." In J.B. Lansing, et al., (eds.), Working Papers on Survey Research in Poverty Areas, Chapter 11. Ann Arbor, MI: Survey Research Center, The University of Michigan.

Courtenay, G. (1978). "Questionnaire Construction." In Hoinville, G., and Jowell, R., Survey Research Practice, chapter 3. Heinemann Educational Books: London.

DeMaio, T., Mathiowetz, N., Rothgeb, J., Beach, M.E., and Durant, S. (1993). "Protocol For Pretesting Demographic Surveys at the Census Bureau." Report of the Pretesting Committee. (U.S. Bureau of the Census).

Esposito, J.L., Rothgeb, J.M. and Campanelli, P.C. (1994) "The Utility and Flexibility of Behavior Coding as a Method For Evaluating Questionnaires." Paper Presented at the annual conference of the American Association for Public Opinion Research, Danvers, MA.

Fowler, F.Jr., (1989). "The Significance of Unclear Questions." In Cannell, C., Oksenberg, L., Kalton, G., Bischoping, K. and Fowler, F.Jr. "New techniques for Pretesting Survey Questions." Research Report. Survey Research Center, The University of Michigan.

Morton-Williams, J., (1979). "The Use of 'Verbal Interaction Coding' For Evaluating a Questionnaire." Quality and Quantity, 13, pp. 59 -75.

Presser and Blair (1994) "Survey pretesting: do different methods produce different results?" In Sociological Methodology, chapter 2.

Sheatsley, P.B., (1983). "Questionnaire Construction and Item Writing." In Rossi, P.H., Wright, J.D., Anderson, A.B. (eds.) Handbook of Survey Research, chapter 6. Academic Press, Inc.: San Diego, CA.

Sudman, S., (1983). "Applied Sampling." In Rossi, P.H., Wright, J.D., Anderson, A.B. (eds.) Handbook of Survey Research, chapter 5. Academic Press, Inc.: San Diego, CA.

Weinberg, E., (1983). "Data Collection: Planning and Management." In Rossi, P.H., Wright, J.D., Anderson, A.B. (eds.) Handbook of Survey Research, chapter 9. Academic Press, Inc.: San Diego, CA.


## Notes

1. The views expressed in this paper are the authors', and do not necessarily represent the official views or positions of the U.S. Bureau of the Census.

2. Eleven of the sample addresses were vacant; we excluded from the analysis the special interviews AHS conducts on vacant units. Other special circumstances resulted in the loss of another 11 interviews (non-sample addresses; "usual residence elsewhere"; interviews not conducted in English; telephone interviews; non-taped observation interviews), and 8 interviews were not taped due to interviewer error and equipment failures. Four respondents refused taping due to a language barrier, and 35 respondents initially refused to be interviewed -- no tape recording was attempted on those which were later converted to complete interviews.

Unfortunately, our experience mirrors that of several other recent Census Bureau behavior coding studies which have also been plagued by high taping failure rates. Although our use of the method in this instance is only to identify potential questionnaire problems, and not to draw conclusions about some larger population, the high level of attrition does raise additional concerns about the validity of our findings, above and beyond those inherent in any qualitative research using non-representative samples.

3. Final interviewer codes included Exact Reading- read as worded or verified correctly; Slight Change- wording was altered but intent of the question was not changed; Major Change- interviewer changed intent of the question; and Omission- interviewer did not read question. Final respondent codes included Adequate- codeable answer; Inadequate- answer which was not codeable; Interrupt- respondent interrupted the interviewer; Qualified- respondent indicated uncertainty with the response; and Request for Clarification- more information about the question was requested.

4. The first 15 interviews came from 2 regional offices and were gathered by a total of 7 interviewers. This subsample included 4 of the 8 types of heating equipment we were most interested in taping. As an aside, this subsample is similar to a pretest one would conduct on a smaller sample size (e.g. 7 interviews from 2 areas versus 16 interviews from 5 areas).

**TABLE 1: AHS Data - Interviewer Codes**

| Qx | N | Exact | Slight | Major | Omission |
|---|---|---|---|---|---|
| **38a** | 15 | 27 | 47 | **27** | - |
| | 31 | 55 | 23 | **23** | - |
| **38c** | 13 | 38 | 46 | **15** | - |
| | 29 | 59 | 34 | 7 | - |
| 38d | 15 | 80 | 7 | 7 | 7 |
| | 31 | 77 | 13 | 6 | 3 |
| **39a** | 15 | 60 | 13 | **27** | -- |
| | 31 | 52 | 23 | **26** | -- |
| **40a** | 15 | 40 | 33 | **27** | -- |
| | 31 | 39 | 42 | **19** | -- |
| 40b | 14 | 50 | 50 | -- | -- |
| | 26 | 58 | 42 | -- | -- |
| **41a** | 15 | 47 | 13 | **33** | 7 |
| | 31 | 48 | 19 | **29** | 3 |
| 42a | 15 | 67 | 27 | -- | 7 |
| | 31 | 65 | 32 | -- | 3 |
| 43a | 15 | 53 | 33 | 13 | -- |
| | 31 | 42 | 52 | 6 | -- |
| 43b | 15 | 47 | 40 | 7 | 7 |
| | 30 | 43 | 43 | 10 | 3 |
| **44** | 15 | 20 | 53 | **20** | 7 |
| | 31 | 16 | 68 | **13** | 3 |
| **45a** | 15 | 13 | 13 | **67** | 7 |
| | 31 | 10 | 19 | **68** | 3 |
| **46a** | 15 | 13 | 33 | **40** | **13** |
| | 30 | 23 | 27 | **43** | 7 |
| **47a** | 15 | 27 | 27 | **47** | -- |
| | 30 | 27 | 27 | **37** | -- |
| 48a | 15 | 7 | 87 | 7 | -- |
| | 31 | 16 | 81 | 3 | -- |
| 48b | 15 | 27 | 67 | -- | 7 |
| | 31 | 32 | 58 | 6 | 3 |
| **48c** | 15 | 20 | 60 | **13** | 7 |
| | 31 | 32 | 55 | 10 | 3 |
| 48d | 15 | 27 | 60 | 7 | 7 |
| | 31 | 23 | 68 | 6 | 3 |
| 48e | 15 | 73 | 13 | 7 | 7 |
| | 31 | 74 | 13 | 10 | 3 |
| 49 | 15 | 20 | 73 | 7 | -- |
| | 31 | 19 | 71 | 10 | -- |
| **50a** | 15 | -- | 13 | **87** | -- |
| | 31 | 10 | 13 | **77** | -- |
| 50b | 15 | 93 | -- | -- | 7 |
| | 31 | 87 | 10 | -- | 3 |

Numbers are in percents

**TABLE 2: AHS Data - Respondent Codes**

| Qx | N | Adequate | Inadequate | Interrupt | Qualified | R Clarify |
|---|---|---|---|---|---|---|
| 38a | 15 | 93 | -- | -- | -- | 7 |
| | 31 | 90 | -- | 6 | -- | 3 |
| 38c | 13 | 92 | 8 | -- | -- | -- |
| | 28 | 89 | 4 | 4 | 4 | -- |
| 38d | 14 | 93 | -- | 7 | -- | -- |
| | 29 | 90 | 7 | 3 | -- | -- |
| 39a | 15 | 100 | -- | -- | -- | -- |
| | 31 | 97 | -- | 3 | -- | -- |
| 40a | 15 | 93 | -- | 7 | -- | -- |
| | 31 | 94 | -- | 6 | -- | -- |
| **40b** | 14 | 79 | -- | **14** | 7 | -- |
| | 26 | 69 | -- | **19** | 12 | -- |
| 41a | 14 | 93 | -- | -- | -- | 7 |
| | 30 | 90 | -- | 3 | -- | 7 |
| 42a | 14 | 93 | -- | -- | -- | 7 |
| | 30 | 93 | -- | -- | -- | 7 |
| **43a** | 15 | 80 | -- | 7 | 7 | 7 |
| | 31 | 84 | -- | 6 | 6 | 3 |
| **43b** | 13 | 77 | 8 | 15 | -- | -- |
| | 27 | 78 | 7 | 11 | -- | 4 |
| 44 | 13 | 100 | -- | -- | -- | -- |
| | 29 | 97 | -- | 3 | -- | -- |
| **45a** | 14 | 43 | 43 | 7 | -- | 7 |
| | 29 | 34 | 41 | 14 | -- | 10 |
| **46a** | 12 | 75 | 17 | 8 | -- | -- |
| | 26 | 73 | 19 | 4 | -- | 4 |
| **47a** | 15 | 67 | 7 | 27 | -- | -- |
| | 30 | 73 | 10 | 13 | 3 | -- |
| **48a** | 15 | 80 | 13 | -- | -- | 7 |
| | 31 | 81 | 6 | 3 | 6 | 3 |
| 48b | 14 | 93 | -- | 7 | -- | -- |
| | 29 | 86 | -- | 3 | 7 | 3 |
| 48c | 14 | 100 | -- | -- | -- | -- |
| | 30 | 100 | -- | -- | -- | -- |
| **48d** | 13 | 54 | 8 | 23 | 8 | 8 |
| | 29 | 66 | 3 | 24 | 3 | 3 |
| **48e** | 14 | 64 | -- | 29 | -- | 7 |
| | 30 | 73 | 3 | 20 | -- | 3 |
| **49** | 15 | 60 | -- | 13 | 27 | -- |
| | 31 | 58 | 3 | 10 | 26 | 3 |
| **50a** | 15 | 47 | 7 | 27 | 20 | -- |
| | 31 | 45 | 3 | 19 | 26 | 6 |
| 50b | 14 | 93 | -- | -- | 7 | -- |
| | 31 | 97 | -- | -- | 3 | -- |

Numbers are in percents