Development and Testing of Business Survey Questionnaires: Use of Qualitative Methods in National Statistical Institutes

Danna L. Moore, Mojca Bavdaž, Deirdre Giesen, Paul A. Smith, Jacqui Jones[1]

Abstract

Qualitative research methods are regarded as an important tool to develop and test survey questionnaires, as they provide useful information for identifying and understanding errors in the response process. However, little is known about how these methods are implemented in practice, especially in the context of establishment surveys. This paper aims to give more insight into whether and how statistical agencies use qualitative methods for the development of establishment survey questionnaires. The paper presents the results of an international survey of national statistical institutes. The survey addressed several aspects of the use of qualitative methods for developing establishment survey questionnaires, including sampling and recruitment procedures, methods for the collection and analysis of the data, and documentation of the results.

Key Words: establishment survey, cognitive interviewing, pretest, questionnaire design

1. Introduction and Background

In survey methodology literature, testing a questionnaire is considered indispensable (e.g. Bradburn, Sudman, and Wansink, 2004). This requirement is echoed in the European Statistics Code of Practice (2011), where one of the indicators for appropriate statistical procedures is that "questionnaires are systematically tested prior to the data collection". Similarly, the Standards and Guidelines for Statistical Surveys of the US Office of Management and Budget (2006) define as a standard that federal agencies must ensure "that all components of a survey function as intended when implemented in the full-scale survey and that measurement error is controlled by conducting a pretest of the survey components or by having successfully fielded the survey components on a previous occasion". One of the guidelines to achieve this standard is to test "new components of a survey using methods such as cognitive testing, focus groups, and usability testing, prior to a field test of the survey system and incorporate the results from these tests into the final design".

National statistical institutes (NSIs) typically publish little about how questionnaires are developed, especially in the context of establishment surveys. By definition, an NSI is the

[1] Danna L. Moore, Washington State University, Faculty Social and Economic Sciences Research Center, PO Box 1801, Pullman, WA 99164-4014 U.S.A., email: moored@wsu.edu. Mojca Bavdaž, University of Ljubljana, Faculty of Economics, Kardeljeva pl. 17, SI-1000 Ljubljana, Slovenia, email: mojca.bavdaz@ef.uni-lj.si. Deirdre Giesen, Statistics Netherlands, PO Box 4481, 6401 CZ Heerlen,The Netherlands, email: d.giesen@cbs.nl. Paul A. Smith, University of Southampton, Highfield, Southampton, SO17 1BJ, U.K., email: p.a.smith@soton.ac.uk. Jacqui Jones, Australian Bureau of Statistics, Australian Bureau of Statistics, 45 Benjamin Way, Belconnen, Canberra ACT 2627 Australia, email: jacqui.jones@abs.gov.au. The views expressed in this paper are those of the authors and do not necessarily reflect the policies of their employers

leading statistical agency within a national statistical system (OECD, 2004). Globally, foundation and responsibilities of NSIs are similar across countries – to collect data to measure economic activity that is vital to countries for economic stability and functioning. NSIs, by their nature, are leaders of business data collection in their respective countries and therefore are a good yardstick of applied research methods in practice.

Various methods can be used to test survey questionnaires. Brancato *et al*. (2006) say that a minimum recommendation is that the questionnaire should be tested at least once by contacting potential respondents. Qualitative methods are often advocated to get input from (potential) respondents; frequently described methods are semi-structured interviews in early scoping studies, for example to assess data availability, focus group interviews, to discuss concept questions or procedures and cognitive test interviews to test concept questions. Cognitive interviewing is seen as an important tool for identifying data errors in the response process, their causes and effects on answering questions about businesses and business activities (Willis 2005). Because the business survey response process is more complex, cognitive interviewing often takes on aspects of exploratory studies as well, resulting in a cognitive hybrid that explores data availability and respondent roles along with cognitive response processes (Willimack 2013).

Little is known about whether and how NSIs implement qualitative testing of business and establishment survey questionnaires, especially cognitive interviewing. We follow in the footsteps of Boeije and Willis (2013), who identified deficiencies in the reporting of results from cognitive interviewing and then went on to test if a framework could be useful to guide reporting. In this effort the authors offered that current cognitive interviewing practice would benefit from the inclusion of 10 categories of information, the Cognitive Interviewing Reporting Framework (CIRF), to guide and systemize information provided in reports. Their objective was to create a reporting standard for cognitive testing studies. In our study, we set out to understand the current state of qualitative questionnaire testing in NSIs. We seek to extend Boeije and Willis's research to more fully understand the broader context of qualitative testing in business and establishment surveys, documenting and describing current applications in NSIs.

Our goal is to provide a resource and empirical data to NSIs to be utilized and referred to whether starting with or furthering their own testing efforts in the development of business survey questionnaires. The insights should be helpful to see if certain methodologies can be considered standards and may help NSIs to benchmark their current practices. For NSIs that for various reasons have not used qualitative testing for their business surveys, this paper may serve as an opportunity to reconsider their testing methods.

2. International Survey of Qualitative Testing Practice for Business and Establishment Surveys

2.1 Survey and Questionnaire Design
To understand the use of qualitative testing methods for question and questionnaire evaluation in business and establishment surveys conducted by NSIs we carried out an international web survey with email invitation of NSIs. The list of NSI (country, NSI name, the director's name and email) was retrieved from the website of the International Statistical Institute (2016). Our list excluded institutions that were not NSIs (e.g. societies and research centers) and those operating at a lower hierarchical level in the national statistical system. One institution was kept per country with the exception of the U.S.A.

where 18 federal agencies were taken into account (because of the decentralized system for business surveys). The NSIs were then assigned to six geographical regions as defined by the United Nations (2016). Our population thus initially consisted of 232 NSIs from 215 countries (see Table 1).

The email invitation was sent either to the director (general) of the NSI or to the person identified as knowledgeable of questionnaire testing. Specific people were mainly identified in European and North American NSIs through the authors' personal links and/or pre-contacts by email to NSIs. When no email was available or only a general one, an attempt was made to find the director's email. This exercise revealed some outdated information (e.g. changes of director) and use of private emails for work purposes in some regions.

Table 1: Geographical distribution of NSIs in our initial population

Geographical region	Countries	NSIs
Africa	53	53
Asia	51	51
Europe	45	45
Latin America and the Caribbean	40	40
Northern America	3	20
Oceania	23	23
Total	215	232

The email invitation explained the purpose of the study and asked for help identifying the best respondent if the recipient did not have enough knowledge of questionnaire testing. The text included the web link to the web survey and an individualized access code. Respondents could also register for entry into the survey. Email and telephone contact was provided in case of questions. The invitation was signed by the international team of five researchers (the authors of this paper).

The questionnaire was drafted, discussed and revised in several iterative steps to first reach a minimum consensus within the international team of five researchers. Questionnaire testing was conducted along with the technical implementation. One NSI completed the questionnaire as part of the pretest before the survey went to field.

The questionnaire addressed five themes: (i) data to determine the eligibility of the unit and the appropriateness of respondent selection; (ii) selection and recruiting for qualitative testing; (iii) design and collection of qualitative interview data; (iv) analysis and reporting of qualitative testing; and (v) an important recent qualitative testing project. Most questions referred to the last five years.

2.2 Implementation and Response
Social and Economic Sciences Research Center, an academic survey center at Washington State University, implemented the survey. The survey was in the field between the end of April and the end of August though most responses came in by the end of June after three email reminders over six weeks. 49 emails had to be resent to new addresses, and even these emails did not reach 14 NSIs, so alternative email addresses were sought. Additional

efforts were made as part of the non-response follow up: sending a personalized email request through a connection if available and asking just three questions in the email to better understand the situation in nonresponding NSIs. These three questions asked whether or not the NSI conducted business surveys, about how many business surveys they conducted annually and whether or not they interviewed or otherwise contacted people from businesses when preparing new or changing old survey questions and questionnaires.

Table 2: Response and Main Characteristics by Geographical Region

Geographical Region	NSIs	Contact Established	Business Surveys Conducted	Qualitative Testing Conducted	Response
Africa	53	15	15	8	3
Asia	51	17	17	9	1
Europe	45	33	33	19	14
Latin America and the Caribbean	40	9	9	3	3
Northern America	20	15	12	11	9
Oceania	23	6	4	3	2
Total	*232*	*95*	*90*	*53*	*32*

Table 2 provides an overview of response by six geographical regions. We managed to establish a contact and get at least some data from 95 or 41% of NSIs. From these contacts we learned that 5 NSIs do not conduct business surveys so they should not have been included in our population of NSIs. Out of these 90 NSIs, 53 or 59% conducted (at least some sort of) qualitative testing of business surveys.

After careful examination of individual answers, 32 NSIs were included in analyses, of which six can be treated as partial respondents. Several NSIs were excluded because their answers had too many missing values or several answers suggested miscomprehension of qualitative research vocabulary (e.g. referring to codes of standard classifications when speaking about coding of data from qualitative interviews). Some NSIs also started completing the questionnaire only to realize that they do not conduct this kind of testing.

2.3 Respondents
Most respondents to the survey described themselves as knowledgeable of qualitative testing of business surveys across their organization. 59% said they knew about qualitative testing of most or all of the business surveys conducted by the organization, 34% of some of the business surveys, and only 6% reporting knowledge of one business survey. More than half, 58% of NSIs had a central team or unit that was responsible for carrying out qualitative research or testing of business questionnaires, many of which had conducted this type of testing in the last two years.

3. Results

Results are based on response from 32 NSIs that provided enough data for a good overview of their practices in qualitative testing studies from start to finish. Results based on questions with some item non-response are calculated from valid answers only; n=32 in the figures unless otherwise specified. The section follows the usual phases of qualitative

testing studies: sampling and recruitment, design and data collection, and analysis and reporting.

3.1 Sampling and Recruitment in Qualitative Testing Studies

Sampling and recruitment are the first steps in a qualitative study. Because of the type of detailed information collected, qualitative studies typically have relatively few cases. There is often some information on the types of cases which can be recruited, but it is not clear how to use this information to best effect in designing a qualitative study. Tourangeau (2004) reviews experimental design approaches to qualitative testing, but practical considerations often have a strong influence on what is actually done. Here we examine the ways in which responding NSIs approach sampling and recruitment.
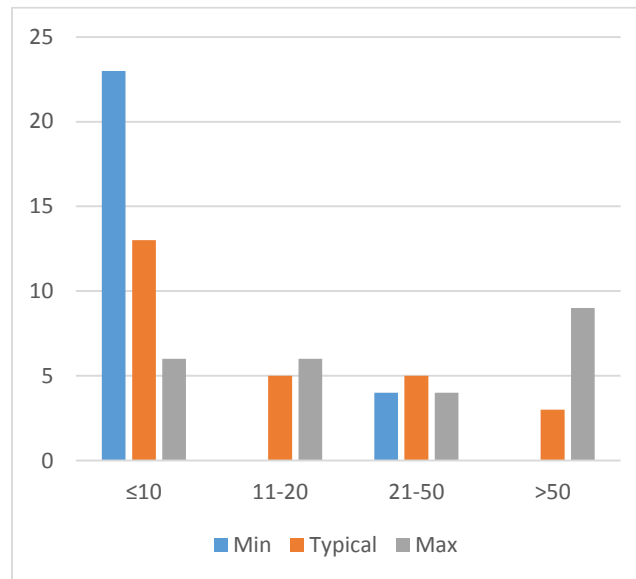
Figure 1: Frequency of Minimum, Typical and Maximum Study Sizes among Respondents (n = 25-27).

One component is how many businesses can be included in a qualitative study. Figure 1 shows the frequencies of different study sizes, specified by respondents as the minimum, typical and maximum sizes. A selection of factors which influence the choice of study size is shown in Figure 2. Resource constraints (budgetary and in terms of specialist staff time) are the commonest "Often/Always" important factors, and agency rules specified a size in some cases. Difficulty in obtaining recruits was "Often/Always" a factor in more organisations than gaining cooperation, but the accuracy of the frame was mostly not an important constraint. The cost of cash incentives was likewise not a common constraint, and indeed 89% of respondent NSIs "Rarely/Never" used incentives.

The resource constraints are also reflected in the sampling approach, where the procedures which are "Often/Always" used in most respondent NSIs are quota sampling (60% of respondents), and sampling based on practical considerations (purposive sampling, for example choosing businesses near to the location of the office to reduce travel costs for specialist staff) (69%).
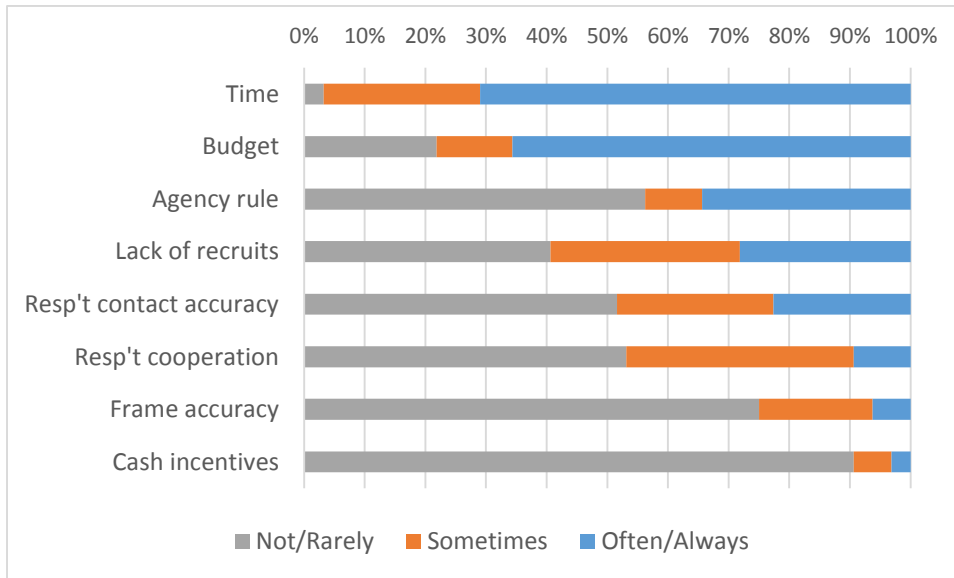
Figure 2: Reasons for Sample Size Choices in Qualitative Research.

Participants are most frequently chosen for qualitative studies because they are already participating in surveys, and almost as frequently because they have previously participated (Figure 3). Recruits are "Often/Always" drawn from businesses which agreed to be recontacted in fewer NSIs. Contacts with businesses are most often with a named survey contact, and there was a very clear hierarchy of contact modes, with 72% of NSIs "Often/Always" using telephone, 48% using email and only 14% "Often/Always" using post.
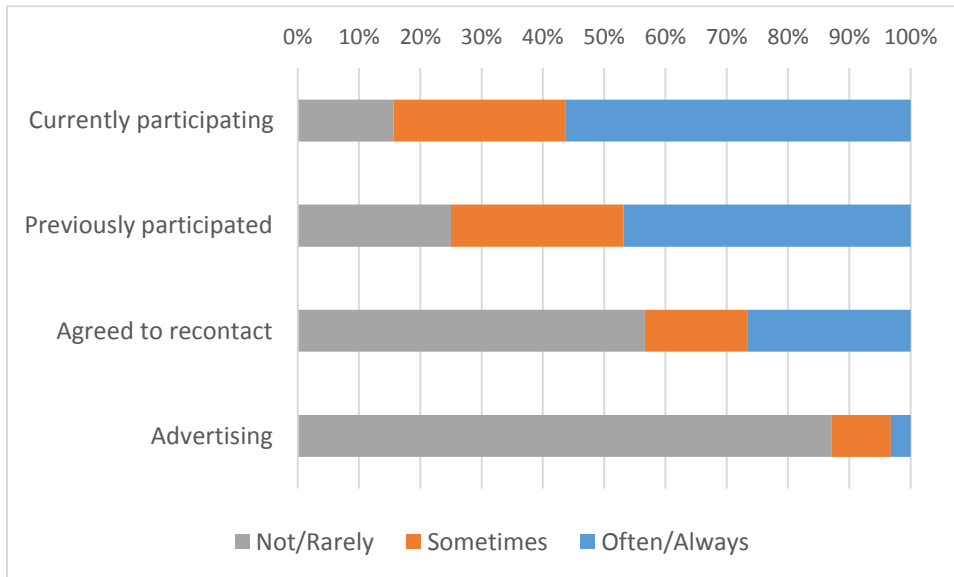
Figure 3: Criteria for Choosing Study Participants.

The experimental design approach (Tourangeau 2005) suggests that a sample should be chosen to cover a range of characteristics. Figure 4 shows that covering a range of sizes is the most frequently used criterion among respondent NSIs, with industrial classification also frequently used. Multilocation sites are also "Often/Always" targeted, which likely

reflects their additional complexity and the need to ensure that questionnaires work for these businesses, which are often the most important for published estimates.
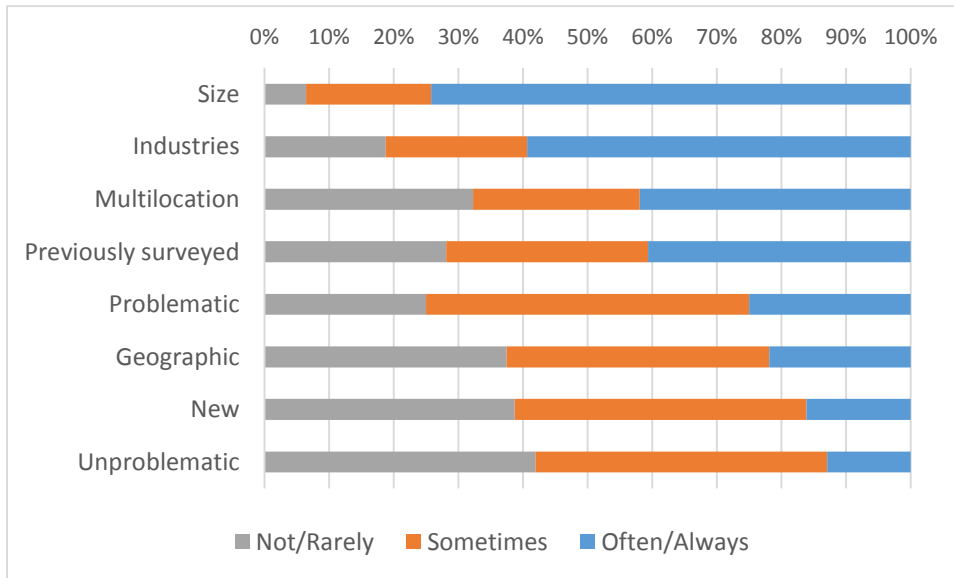

Figure 4: Range of Characteristics of Businesses to be Covered within Samples in Qualitative Studies.

3.2 Design and Data Collection in Qualitative Testing Studies

Qualitative testing and research are often used in business surveys to develop and refine survey questionnaires and other instruments for use with business respondents in terms of questionnaire content, measurements collected, and survey questions. One of our expectations in this study is that NSI business survey development and qualitative question testing, overall, has describable features and it should be possible to describe why particular variants of qualitative testing are used in practice.
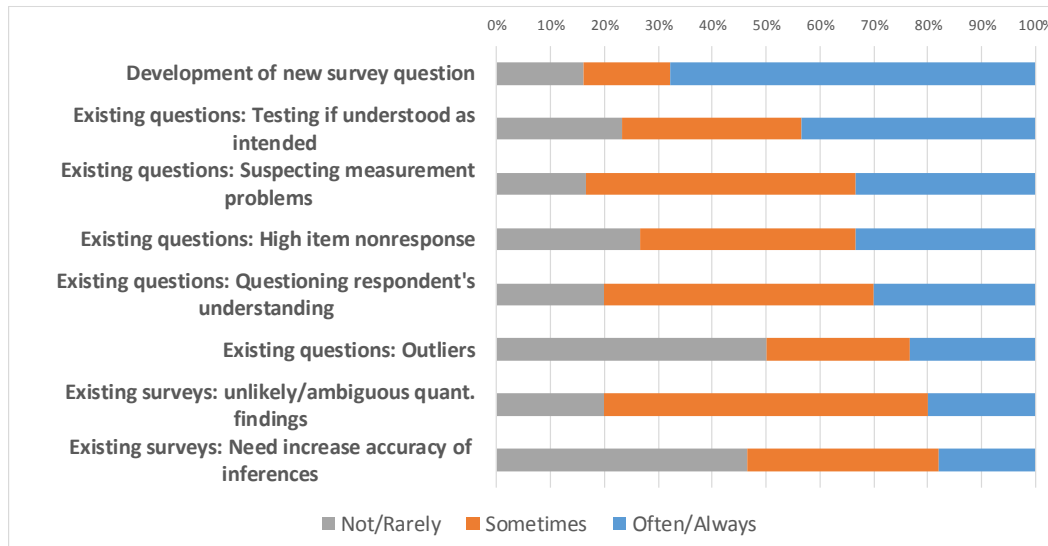

Figure 5: Reasons for Conducting Qualitative Research in Business Surveys (n = 28-31).

Figure 5 thus displays the driving issues for conducting qualitative research and testing for business questionnaires and shows that the leading "Often/Always" reasons were: the development of new survey questions (68%), to test if existing questions were understood as intended (43%), the need to increase accuracy of existing survey questions with suspect measurement problems (33%) and high item nonresponse.

We are interested in documenting the types and methods of qualitative testing that are performed and used by NSI's for business surveys.

The procedures and methods generally used to conduct qualitative research of existing or new survey questions most frequently reported are displayed in Figure 6. The most "Often/Always" used procedures were: cognitive interviews (55%), pretest interviews (52%), and usability tests (45%). The least used "Often/Always" method was focus groups (23%).

Conduct of qualitative research is handled "often/always" by staff with specialized knowledge and experience in qualitative research (73%), staff who drafted and developed the questions (61%), staff with at least a Masters degree (50%) and staff from the business area with statistical or content expertise (43%). Only 27% reports that the qualitative research is "often/always" conducted by interviewers or field staff. When interviews are conducted, most often this involves 1-2 respondents from the business to be interviewed, and 2 staff from the NSI attending interviews. Generally, 2-3 staff of an NSI are involved in the conduct of the project to research questions.
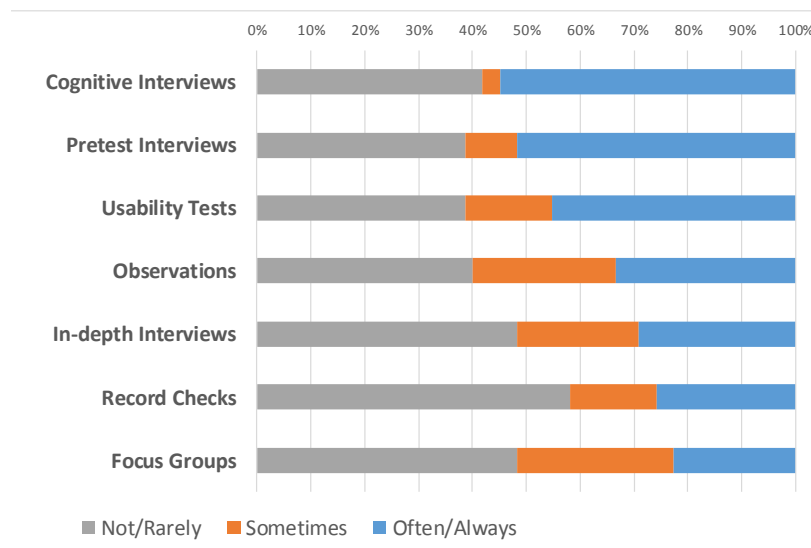

Figure 6: Frequency of Qualitative Research Procedures in Business Surveys (n = 30-31).

The process of qualitative testing with businesses involves a number of tasks and many approaches can be used. The predominant procedures "Often/Always" used by the majority of NSIs was to prepare topics to discuss during the interview (84%) and to prepare probes that might be used during the interview (79%). Preparing a detailed script to be used verbatim during the interview was used "Often/Always" by 26% of the respondents and sending a questionnaire in advance of the interview by 10%. When asked how often they asked respondents to complete questions or the questionnaire prior to qualitative testing, 61% reported "Not at all/Rarely", 29% "Sometimes" and 10% "Often/Always".

As surveys have evolved over time, the modes of administration have changed, involve multiple survey modes, or have shifted towards new technologies. One of the largest concerns for testing of questions and questionnaires, as more modes are involved, is determining equivalency of measurement and the need to test questions and questionnaires in multiple modes. This current study provides a benchmark of the use of modes during qualitative research and testing.

Figure 7 displays the use of survey modes and other methods associated with business survey qualitative testing. The leading methods used "Often/Always" were: face-to-face interviews (76%) and observations (36%). The remaining modes are used less frequently (less than 20%) on a regular basis as main components in qualitative testing for business surveys. If questionnaires were available in all survey modes, 44% said they "Often/Always" tested and evaluated in all modes, 40% said they do this "Not at all/Rarely" and 16% do this sometimes. Seven responding NSI's said they did not do any mixed mode surveys. The environment predominantly used by organizations for testing is the respondent's workplace or desk at the business (68%), in a meeting room at the business's site (56%). Least often (8%, Often/Always) was asking respondents to come to the organization.
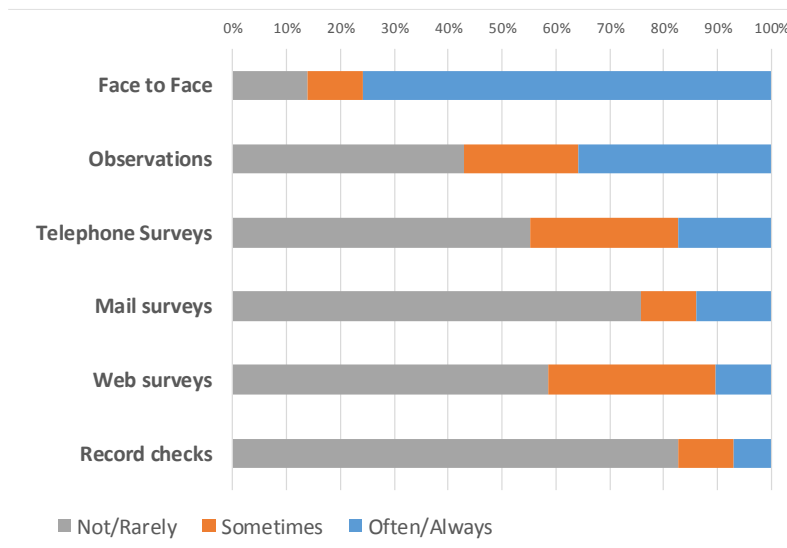

Figure 7: Frequency of Use of Survey Modes and Procedures Qualitative Testing (n = 28-29).

During qualitative testing interviews, three types of probing were used. The most frequently used probe was spontaneous probes (53% use these "Often/Always"), followed closely by retrospective probes and concurrent probes as shown in Figure 8. Of the tasks that business respondents are asked to do during testing, "to answer the interviewer's probes" is the most common with 74% asking for this "Often/Always". This is closely followed with more detailed and comprehensive tasks of respondents: explaining how they completed the questionnaire (67%), to complete the questionnaire (65%), to explain how they would go about completing the questionnaire (60%) and to think aloud as they complete the questionnaire (57%).
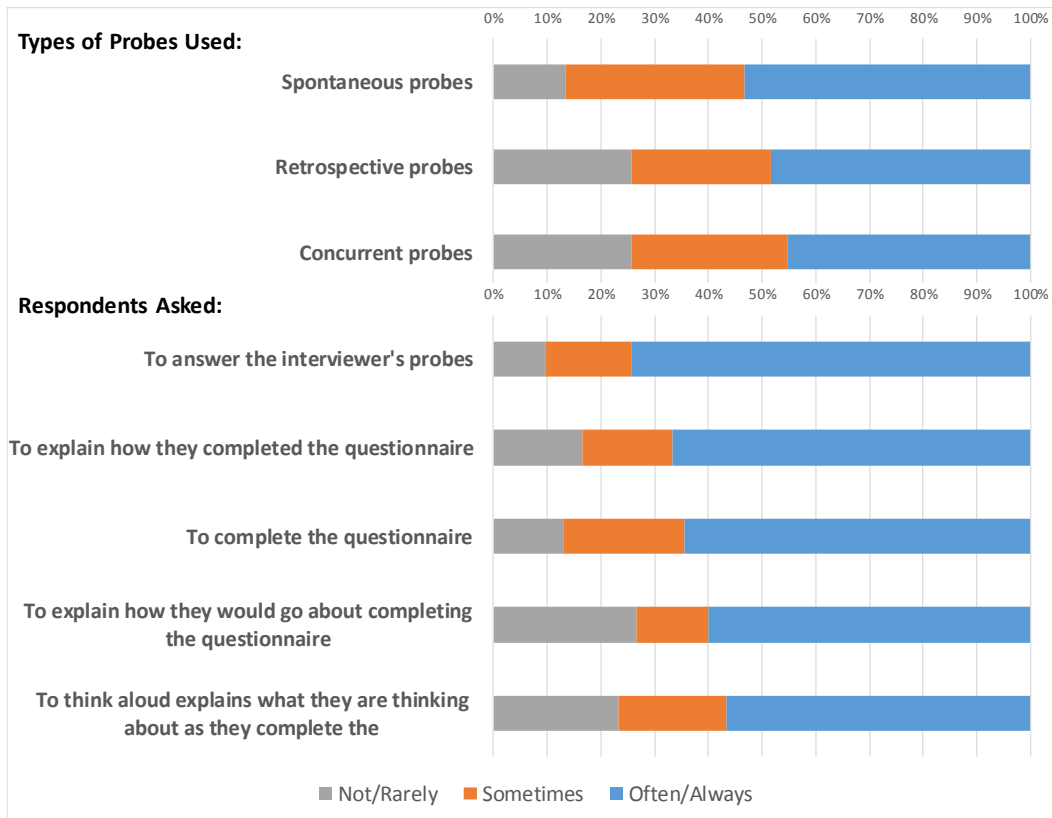
Figure 8: Types of Probes Used and Types of Tasks Respondents Are Asked to Do During Qualitative Testing Interviews (n = 30-31).

For the documentation of qualitative interviews (see Figure 9) 90% of the respondents indicate that they "Often/Always" take notes, audio recording is "Often/Always" made by 38% of the respondents, the same percentage is reported for the collection and entry of completed questionnaires. Less frequently "Often/Always" used are professional on-site transcription (13%), video recording (10%) and eye tracking (4%). Respondents provided examples in open-ended questions about the other ways qualitative interviews were conducted and documented, including: (i) Note write-up and active recall of interviews; (ii) in-situ (site of origin) noting on the paper questionnaire; (iii) careful noting relative to "known problems" and "where it takes place" on the questionnaire; (iv) annotated questionnaires; (v) use of standardized testing reporting templates; (vi) for web surveys, taking notes and marking power point screen shots or web pages for questions; (vii) using Skype remote user testing with shared screens. Some respondents reported what they considered to be "better practices" they have for documenting the respondent experience in usability testing. These included: conducting testing at the business site; conducting testing and interview at the computer of the respondent for the business; during the interview sitting with respondent such that the respondent's screen accessing the survey can be seen and is viewable; adding reminder notes to survey screenshots to prompt note takers to observe for particular respondent behaviors with respect to particular questions.
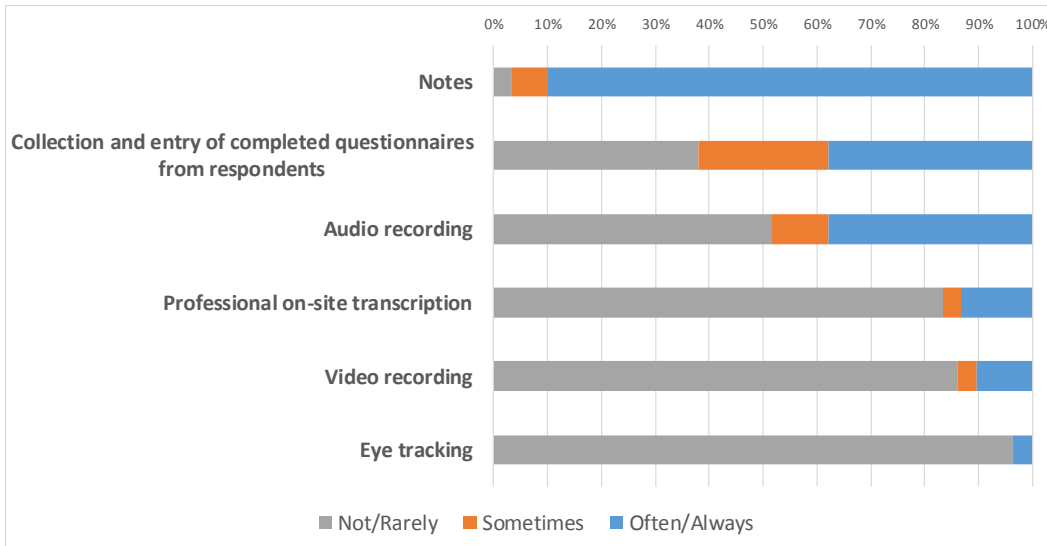
Figure 9: Ways of Documenting Qualitative Interviews (n = 28-30).

3.3 Analyses and Reporting in Qualitative Testing Studies

Best practice for analysis in qualitative research is to collect rich data and then analyze them thoroughly with more than one person who reach consensus together. Various forms of shortcuts may be used for various reasons (usually budget, staff, time). The most critical practice is to only rely on a single person who rereads the notes of interviews because so much data is lost by that point. Figure 10 presents practices that are used in analyzing qualitative interview data in NSIs. 83% of respondents indicated that notes are "Often/Always" reread and actually just one respondent claimed they never or rarely reread notes. 61% of respondents "Often/Always" agree on findings from interviews and 57% discuss each reviewer's findings and compare them to others. More than half of respondents (57%) also summarize data from each interview in a standard format. Less frequently (47%) more than one person "Often/Always" analyses the same data. Listening
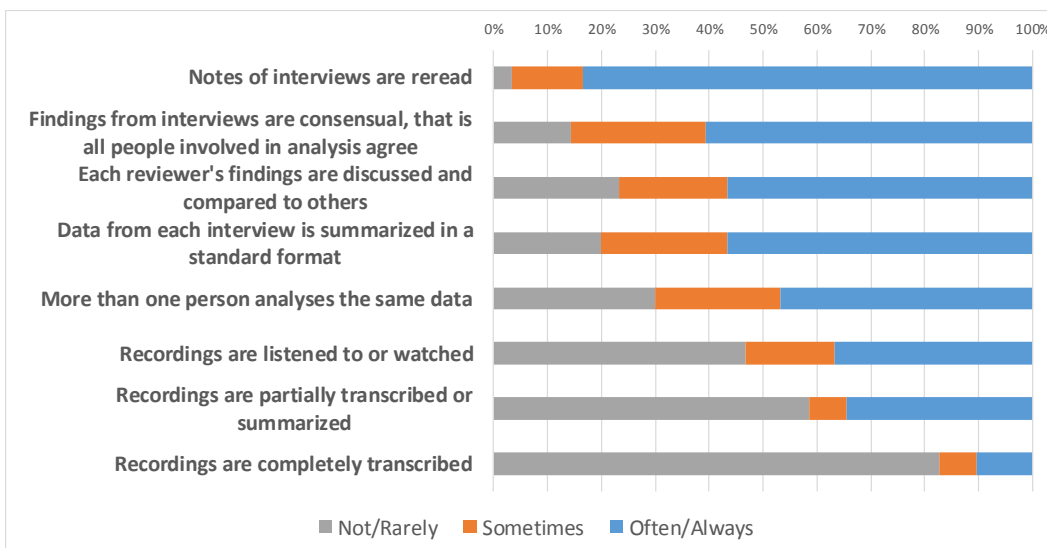

Figure 10: Practices Used in Analyzing Qualitative Interview Data (n = 28-30).

to or watching recordings is even less present, only about a third of respondents do it "Often/Always" (37%). Transcriptions are not common for qualitative research practice of NSIs: about 35% of respondents "Often/Always" partially transcribe or summarize recordings while only 10% "Often/Always" completely transcribe them.

Figure 11 shows how qualitative interview data are summarized once data have been collected and analyzed. Summaries at the question level are the expected form as they can be directly used for improving the questions and the questionnaire, and also allow to use this knowledge for later surveys. 80% of respondents "Often/Always" summarize data this way. Other ways of summarizing data offer other insights and benefits but are somewhat less widespread. 57% of respondents "Often/Always" prepare summaries by types of errors, issues and problems, thus possibly contributing to knowledge on questionnaire design. Less than a third of respondents (31% and 27% respectively) "Often/Always" prepare summaries by industry and size that might inform sampling and subject-matter specialists. Only a fifth of respondents "Often/Always" prepare summaries by multi sites vs. single location.
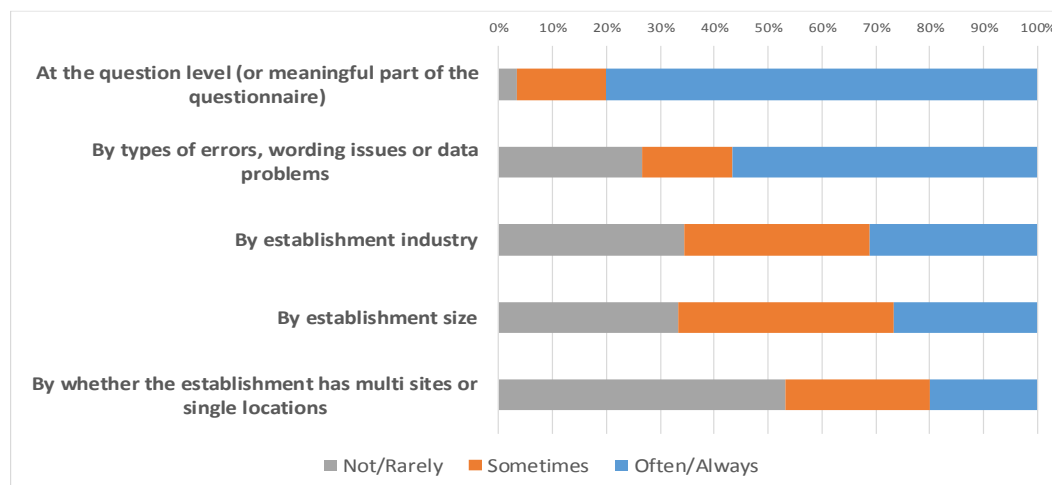

Figure 11: Ways of Summarizing Qualitative Interview Data (n = 29-30).

A typical way of documenting qualitative interview studies is in a report. 77% of respondents "Often/Always" prepare a report. About two third (65%) also "Often/Always" present results in an internal meeting while presenting results outside the organization is less common.

4. Discussion and Conclusion

Qualitative research methods are regarded as an important tool to develop and test survey questionnaires, as they provide useful information for identifying and understanding errors in the response process. This paper presents the first results of an international survey of NSIs on qualitative testing practices in the context of business and establishment surveys. While great effort was invested in reaching as many NSIs from different parts of the World as possible, the number of NSIs revealing their qualitative testing practices to us is not high (the figures are still considered preliminary as some NSIs recently expressed their willingness to contribute more data). One reason is that qualitative testing practices for business surveys are not globally present. Another reason is that many NSIs do not have a central unit for testing questionnaires and that the task of testing is left to the discretion of

each survey manager. An important fact also seems to be that vocabulary to discuss qualitative testing practices is still developing and/or is not widely known among staff without a qualitative research background. Therefore, response to our survey might be highly selective, with a higher likelihood of response for NSIs familiar with qualitative testing. Despite the low number of NSIs included in our analysis and no guarantee that results are representative, we seem to get a fairly good insight into the practices of NSIs that do use qualitative testing.

The results suggest that there are some deviations in the qualitative testing practices of NSIs from the best practices of qualitative research. However, about half of respondents said that they at least sometimes run into problems that may limit the extent and/or depth of qualitative testing studies, for instance, gaining cooperation from the largest businesses or across all businesses, and analyzing data more systematically or more in depth. Actually, it seems much more problematic that some NSIs do not do these kinds of studies on establishment surveys at all. Those that do qualitative testing studies still follow many good practices of qualitative research but there is potential, for instance, to improve data capture and the richness of data to be analyzed, and to benefit from transcriptions to support findings and recommendations. Future research should provide more understanding of how much specific practices of qualitative testing contribute to the quality of business surveys.

Acknowledgements

References

Boeije, H., and Willis, G. (2013). The Cognitive Interviewing Reporting Framework (CIRF): Towards the harmonization of cognitive interviewing reports. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 9(3),* 87-95.

Bradburn, N., Sudman, S., and Wansink, B. (2004). "Asking Questions" (Rev. ed.). San Francisco, CA: Jossey-Bass.

Brancato, G., Macchia, S., Murgia, M., Signore, M., Simeoni, G., Blanke, K., Körner, T., Nimmergut, A., Lima, P., Paulino, R., and Hoffmeyer-Zlotnik, J. (2006). "Handbook of Recommended Practices for Questionnaire Development and Testing in the European Statistical System". Retrieved from http://ec.europa.eu/eurostat/ramon/statmanuals/files/Handbook_of_Practices_for_Quest.pdf

European Commission (2011). "European statistics code of practice for the national and community statistical authorities". Adopted by the European Statistical System Committee 28th September 2011. Retrieved from http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-32-11-955

International Statistical Institute. (n.d.). "National Statistical Offices". Retrieved March 11, 2016 from https://www.isi-web.org/index.php/resources/national-statistical-offices

OECD (2004, February 4). "National Statistical Office (NSO)". Glossary of Statistical Terms. Retrieved from https://stats.oecd.org/glossary/detail.asp?ID=4344

Tourangeau, R. (2004) "Experimental design considerations for testing and evaluating questionnaires". In Presser, S., Rothgeb, J.M., Couper, M.P., Lessler, J.T., Martin, E., Martin, J. and Singer, E. (Eds) Methods for Testing and Evaluating Survey Questionnaires (pp. 209-224). Hoboken, NJ: Wiley.

U.S. Office of Management and Budget (2006). "The Standards and Guidelines for Statistical Surveys". Retrieved from https://www.whitehouse.gov/sites/default/files/omb/inforeg/statpolicy/standards_stat_surveys.pdf

United Nations. (2016). "Composition of macro geographical (continental) regions, geographical sub-regions, and selected economic and other groupings". Retrieved June 6, 2016 from http://unstats.un.org/unsd/methods/m49/m49regin.htm

Willimack, D. (2013). "Methods for the development, testing, and evaluation of Data Collection Instruments" In Snijkers, G., Haraldsen, G., Jones, J., and Willimack, D. (Eds.), Designing and Conducting Business Surveys (pp. 253-301). Hoboken, NJ: Wiley.

Willis, G. (2005). *Cognitive interviewing: A tool for improving questionnaire design.* Thousand Oaks, C.A: Sage.