2020

Chapter 4: How to Conduct Effective Interviewer Training: A Meta-Analysis and Systematic Review Appendix 4

Jessica Daikeler

Michael Bosnjak

Chapter 4: How to Conduct Effective Interviewer Training: A Meta-Analysis and Systematic Review

Appendix 4

Jessica Daikeler and Michael Bosnjak

Full book reference:

Olson, K., J. D. Smyth, J. Dykema, A. L. Holbrook, F. Kreuter, B. T. West. 2020. *Interviewer Effects from a Total Survey Error Perspective*. Boca Raton: CRC Press.

Appendix 4A

Appendix Table A4A.1: Overview of the literature on interviewer tasks addressed in interviewer training experiments

Interviewer task	Survey error potentially introduced	Outcomes addressed in interviewer training experiments	References
Generate sampling frame	Coverage error	None	None
Make contact, gain cooperation, gain consent to additional parts of the survey	Unit nonresponse error	Unit nonresponse (response rate)	Basson and Chronister, 2006; Dahlhamer et al., 2010; Cantor et al. 2004; Billiet and Loosveldt 1988; Mayer and O'Brien, 2001; Schnell and Trappman 2006; Durand et al., 2006; Groves and McGonagle 2001; Karlsson 2010
Ask survey questions, record answers, conduct measurements and maintain motivation	Measurement error	Correctly administered, read, probed and recorded items, item nonresponse, accurate responses	Guest, 1954; Benson and Powell, 2015; Dahlhamer et al. 2010; Billiet and Loosveldt 1988; Fowler and Mangione 1986; Cannell et al. 1977; Miller & Cannell 1982
Process the collected data	Processing error	None	None

Appendix 4B The meta-analytical process

This section describes the five steps of the meta-analytic procedure employed in the present study: 1) a comprehensive literature search; 2) checking of the eligibility of studies found; 3) coding of relevant data; 4) calculation of training effect sizes; 5) analysis of variables that moderate effect size (Lipsey and Wilson 2001; Borenstein et al. 2009)

4B.1 Eligibility criteria and search strategy

One of the first steps in a meta-analysis is the definition of the criteria that studies must meet if they are to be included. Table A4B.1 lists these eligibility criteria.

To ensure the quality of the meta-analysis, a comprehensive literature search was conducted. Because a meta-analysis that includes only published literature faces the problem of publication bias, grey literature was also eligible for inclusion (for further information, see Table A4B.1). During the search process, the most common reasons for the exclusion of studies were the lack of an experimental design and missing data quality indicators. Most of the studies rated the use of interviewer training as appropriate but did not evaluate how effective it was.

The PRISMA diagram (Moher et al. 2009) in figure 1 gives an overview of the search strategy. The search was limited to literature in English; over 2,000 results had to be excluded because the broad search terms led to literature related to job interviews, linguistic interviews, cognitive and clinical interviews of victims and witnesses, and studies without an experimental setting. Fourteen eligible publications were retrieved. Because many of the publications presented more than one experiment or effect size, the search yielded a total of 68 experimental comparisons. The most common indicator of data quality was the effect of interviewer training on the response rate (22), followed by the effect on correct recording of the response (14); on item nonresponse (12); on the reading of questions exactly as worded (12); on correct probing (6); and on correct item administration (4).

Table A4B.1 Eligibility criteria

Eligibility Criterion	Description
Experimental design	Studies must employ an experimental design. We accepted both treatments versus control and pre-versus post group designs. In the first case, a group of trained interviewers is compared with a group of less trained or untrained interviewers. While in the pre-versus post-design group the experiment has up to four steps. First, the interviewers receive no or only elementary training, in the second step the data quality is measured, then the interviewers receive professional training, and in the fourth step, the data quality is measured again.
Downgraded training for control group	For both types of training, it was essential that the control group received either no or only an introductory briefing.
Data quality measures	Data quality measures indicating the effectiveness of training are mandatory.
Training content on refusal avoidance and/ or measurement-related data quality	The interview tasks can be divided into two main areas. First, to encourage respondents to participate (nonresponse errors) and second, to achieve adequate data quality during the interview (measurement and processing errors). Therefore, the last selection criterion differs according to the interviewer's task and the measured data quality indicator. For the first task, the avoidance of refusals, we include studies with a classical refusal avoidance training (see Groves and McGonagle 2001). For the second task to improve data quality indicators in the survey process, data quality and interviewer behavior had to be an essential part of the training.

4B.2 Coding procedure

Coding was performed by two independent coders (the coding scheme can be found in appendix Table A4B.4). The lead coder coded all studies and instructed the second coder, who coded 30 percent of the studies. Intercoder reliability produced a Krippendorff's alpha (Krippendorff 2004) of .9 for the effect sizes and .95 for the moderator variables, indicating a match of at least 90 percent between the two coders. Reliability values of .8 and above indicate an almost perfect match (Hallgren 2012). Consequently, it can practically be ruled out that the effect sizes and moderator codings on which this meta-analysis is based were subjectively distorted by the coders.

4B.3 Effect size metric and statistical method

During the search process, it became clear that interviewer training experiments report a variety of different data quality indicators as effect size metrics. From a methodological point of view, most of these data quality indicators are not substantively comparable, which is why it was decided to conduct a separate meta-analysis for each indicator (an overview provides Table A4B.2). As the effect size metric was the same for all seven data quality indicators, the effect sizes were calculated as follows (e.g., for correctly administered items):

$$RD = \frac{Ncait}{Nait} - \frac{Ncaiu}{Naiu}$$

with RD = Rate Difference ,

Ncait = Total Number Of Correctly Administered Items For Trained Group,

Nait = Total Number Of Items For Trained Group,

Ncaiu= Total Number Of Correctly Administered Items For Untrained Group,

Naiu = Total Number Of Items For Untrained Group

The statistical analysis for each of the data quality indicators comprised five steps (Lipsey and Wilson 2001). First, the weighted mean response rate difference across all studies was computed. This variance component consisted of the study-level sampling error variance as well as an estimate of between-study variance (Borenstein et al. 2009). A random-effects analysis was used, as inference should be made for a population of studies larger than the set of observed studies (Hedges and Vevea 1998). In the next step, the confidence interval for the mean effect size was determined to indicate the degree of precision of the estimate and whether the mean effect size was statistically significant. In the third step, a homogeneity analysis was performed to assess whether the effect sizes came from the same population (random effects assumption). In the fourth step, the robustness and quality of the findings were checked with an outlier analysis and publication bias checks. In the final step, a mixed-effect model analysis was performed for each moderator variable to determine which variables had a significant influence on the response rate differences. Studies that did not provide information on moderator variables were excluded from the respective analyses. The R package metafor" (version 1.9-9) was used for the analyses (Viechtbauer 2010). However, not all effect sizes could be used in the closer inspection of the effect sizes, so analyses were only possible for three of the six effect sizes.

4B.4 Publication bias and sensitivity analyses

In the next step, we examined whether a publication bias might have affected the estimates of the mean effect size. To this end, we checked both the funnel plots and the Egger's regression tests (see appendix Figure A4B.1 and Table A4B.3) and found that a publication bias problem existed, as a disproportionate number of significant results had been included in the meta-analyses. One reason for this may have been the generally insufficient number of studies in this area. Outlier tests were conducted in the sensitivity analysis. For response rate and item nonresponse, 10 percent of outlier studies were excluded, and no significant difference between the original and outlier-adjusted effect sizes was found.

Table A4B.2 Description of effect sizes

Response Rate	Experimental interviewer group received Refusal-Avoidance-Training (RAT) and control group did not, invited vs. participated respondents in each group.
Item Nonresponse	Experimental interviewer group received advanced interviewer training and control group not, counting item nonresponse in each group.
Response Accuracy	Experimental interviewer group received advanced interviewer training and control group not, counting accurate responses in both groups.
Administering	Experimental interviewer group received advanced interviewer training and control group not, counting correctly administered questions per interview (audio tape error index).
Probing	Experimental interviewer group received advanced interviewer training and control group not, counting correctly probed questions per interview (audio tape).
Reading	Experimental interviewer group received advanced interviewer training and control group not, counting correctly read questions per interview (audio tape).
Recording	Experimental interviewer group received advanced interviewer training and control group not, counting correctly recorded questions per interview (audio tape).
Reporting	Experimental interviewer group received advanced interviewer training and control group not, comparing reportings on sensitive and exact-reporting items

4B.5 Publication bias

Publication bias exists if the preparation, submission or publication of research findings depend on characteristics of just these research results, e. g. their direction or statistical significance. Publishing only results that show a significant finding disturbs the balance of findings (Weiss and Wagner 2011). We used three techniques to overcome this problem. First, we examined conference abstracts (American Association for Public Opinion Research (AAPOR), European Survey Research Association (ESRA), Joint Statistical Meeting (JSM)), second we used the reference lists of the already located manuscripts and applied a snowballing technique and the last strategy was to ask for appropriate research via mailing lists and email. We followed conference presentations and papers with restricted access by email and asked in this regard for similar research.

The funnel plots in Figure A1 are a visual method used to inspect publication biases (Egger et al. 1997). It shows the individual observed effect sizes on the x-axis against the corresponding standard errors. It is important that the point cloud on both sides of the line is approximately equal in number and distribution, which is not for all of our effect sizes the case. These results are emphasized by the Egger's regression test, which tests the asymmetry of the funnel plot (see Appendix Figure A4B.1).

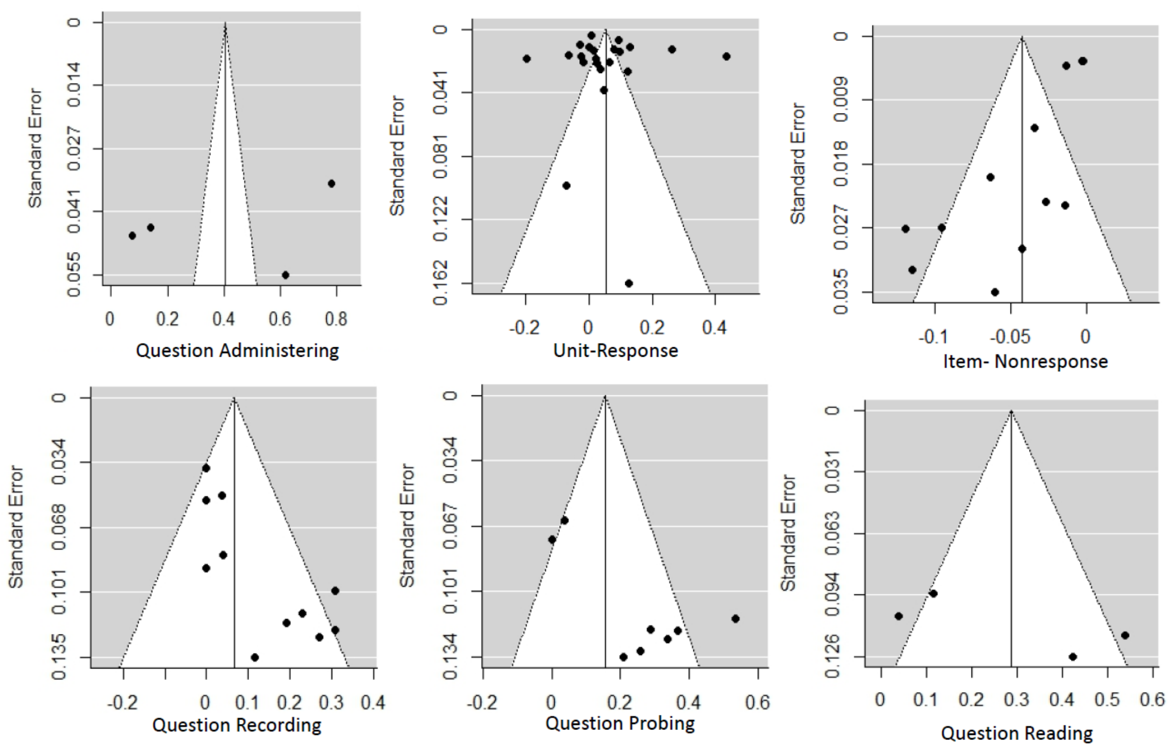

Figure A4B.1: Publication bias: Funnel plots for data quality indicators

Table A4B.3 Publication bias check: Egger's regression test

Effect size measure	Regression Test for Funnel Plot Asymmetry
Response Rate	0.5113
Administration	0.5111
Item Nonresponse	0.0005

Table A4B.4 Coding scheme

Variable	**Scale/Categories**
Case Number	String
Authors	String
Reference	String
Title	String
Year	Continuous
Published	2- Yes/ 1- No
Experiment Number (if study has more than one)	Continuous
Identifier	String
Invited in treated Group	Continuous
Participated in treated Group	Continuous
Number of Interviewer in treated Group	Continuous
Number of Interviews in treated Group	Continuous
Invited in untreated Group	Continuous
Participated in untreated Group	Continuous
Number of Interviewers in untreated Group	Continuous
Number of Interviews in untreated Group	Continuous
Pre/Post or Control/Treatment	2- Control/ Treatment 1- Pre/Post
Control group had also a basic training	2- Yes/ 1- No
Listened to audio refusals	2- Yes/ 1- No
Prior Experiences interviewers	2- Yes/ 1-No
Length of Training in hours	Continuous
Using supplementary Training material	2- Yes/ 1-No
Monitoring	2- Yes/ 1-No
Practice & Feedback Sessions included	2- Yes/ 1-No
Training for Telephone Interviewers only	2- Yes/ 1-No
Training for Face to Face Interviewers only	2- Yes/ 1-No
Includes Blended Learning	2- Yes/ 1-No
Training for all modes	2- Yes/ 1-No
Refusal Avoidance Training Only	2- Yes/ 1-No

Appendix 4C Random effects model and meta regression summary statistics

Table A4C.1 Sampling error weighted mean effect sizes and heterogeneity

Meta-analytic model)	Summary Statistics (random effect			Heterogeneity Estimators		
Data Quality Indicator	K	Mean Response Difference (95% CI)	T (se)	Q_e total (df/p)	I	H
Response Rate	22	0.053 (-0.008/0.1069)	0.0155 (0.0051)	1355.9482 (21/0.0001)	98.96%	94.6. 49%

Appendix 4D List of missing studies in the paper

Missing studies of this systematic overview are Belli and Lepkowski (1996), Marquis (1970), Miller and Cannell. (1977), O'Brien, Mayer, Groves, and O'Neill (2002), Oksenberg, Vinokur, and C. Cannell (1979a), and Oksenberg, Vinokur, and C. Cannell (1979b).

References

Basson, D. and M. Chronister. 2006. Recordings of Prior Refusals: Do They Improve Later Conversion Attempts? *Methodology of Longitudinal Surveys Conference (MOLS 2006)* 12-14 July 2006, Essex, England.

Belli, R. F. and J.M. Lepkowski. 1994.6. Behavior of Survey Actors and The Accuracy of Response. In: Warnecke, R. B. (ed) Health Survey Research Methods: Conference Proceedings (pp. 69-74).

Benson, M. S. and M. B. Powell (2015). Evaluation of a Comprehensive Interactive Training System for Investigative Interviewers of Children. Psychology, Public Policy, and Law 21: 309-322.

Billiet, J. and G. Loosveldt (1988). Improvement of the Quality of Responses to Factual Survey Question by Interviewer Training. Public Opinion Quarterly 52: 190-211.

Borenstein, M., L. V. Hedges, J. Higgins and H. R. Rothstein (2009). Introduction to Meta-Analysis. Hoboken, CA, Wiley Online Library.

Cannell, C. F., L. Oksenberg and J. M. Converse (1977). Striving for Response Accuracy: Experiments in New interviewing Techniques. Journal of Marketing Research 14: 306-315.

Cantor, D., B. Allen, S.J. Schneider, T. Hagerty-Heller, and A. Yuan.2004. Testing an Automated Refusal Avoidance Training Methodology. *Annual Meeting of the American Association for Public Opinion Research (AAPOR)*, Phoenix, AZ.

Dahlhamer, J. M., M. L. Cynamon, J. F. Gentleman, A L. Piani and M. J. Weiler (2010). Minimizing Survey Error through Interviewer Training: New Procedures Applied to the National Health Interview Survey (NHIS).

Durand, C., M.-E. Gagnon, C. Doucet and E. Lacourse (2006). An Inquiry into the Efficacy of a Complementary Training Session for Telephone Survey Interviewers. Bulletin de Méthodologie Sociologique 92: 5-27.

Fowler, F. J. and T. W. Mangione (1986). Interviewer Characteristics And Nonresponse. Proceedings of the 1986 American Statistical Association, Survey Research Methods Section.

Groves, R. and K. A McGonagle (2001). A Theory-Guided Interviewer Training Protocol Regarding Survey Participation. Journal of official statistics VOL 17.

Guest (1954). A New Training Method for Opinion Interviewers. Public Opinion Quarterly 18: 287-299.

Hallgren, K.A. 2012. Computing Inter-Rater Reliability for Observational Data: an Overview and Tutorial. *Tutorials in Quantitative Methods for Psychology*, 8:23.

Hedges, L. V. and J. L. Vevea (1998). Fixed-and Random-Effects Models in Meta-Analysis. Psychological Methods 3: 484.6.

Karlsson, A.Ö. 2010. Recent Developments In Interviewer Training at Statistics Iceland: Minimizing Interviewer Effects and Reducing Refusal Rates. Nordisk Statistikermøde 1, 7.

Krippendorff, K. 2004. Reliability in content analysis: Some common misconceptions and recommendations. *Human communication research* 30(3):411–433.

Marquis, K. 1970. Effects of Social Reinforcement on Health Reporting in the Household Interview. Sociometry. 33:203-215.

Mayer, T.S. and E. O'Brien. 2001. Interviewer Refusal Aversion Training to Increase Survey Participation. In *Proceedings of the Annual Meeting of the American Statistical Association Atlanta, Georgia, August 5 - 9 2001*.

Miller, P. V., and C. Cannell. 1977. Communicating Measurement Objectives in the Interview. In Strategies for Communication Research. Ed. Hirsch, P.M., P. Miller, and F.G. Kline, pp 127-151. Beverly Hills: Sage Publications.

Miller, P. V. and C. F. Cannell (1982). A Study of Experimental Techniques for Telephone Interviewing. Public Opinion Quarterly 46: 250-269.

Moher, D., A. Liberati, J. Tetzlaff and D.G. Altman. 2009. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the Prisma Statement. *Annals of Internal Medicine* 151:264–269.

O'Brien, E., T.S. Mayer, R.M. Groves and G.E. O'Neill. 2002. "Interviewer Training to Increase Survey Participation." Pp. 2502-07 in Survey Research Methods: American Statistical Association. http://www.asasrms.org/Proceedings/y2002f.html

Oksenberg, L., A. Vinokur, and C. Cannell. 1979a. The Effects of Commitment to Being a Good Respondent on Interview Performance. In Experiments in Interviewing Techniques: Field Experiments in Health Reporting, 1971-1977, ed. Cannell, C.L. Oksenberg, and J. Converse. Washington, DC: Department of Health, Education, and Welfare, DHEW Publication No (HRA) 78-3204.

Oksenberg, L., A. Vinokur, and C. Cannell. 1979b. The Effects of Instructions, Commitment and Feedback on Reporting in Personal Interviews. In Experiments in Interviewing Techniques: Field Experiments in Health Reporting, 1971-1977, ed. Cannell, C., L. Oksenberg, and J. Converse. Washington, DC: Department of Health, Education, and Welfare, DHEW Publication No (HRA) 78-3204.

*Schnell, R. and M. Trappman. 2006. The Effect of the Refusal Avoidance Training Experiment on Final Disposition Codes in the German ESS-2. Report, Working Paper 3/2006, Germany: Center for Quantitative Methods and Survey Research, University of Konstanz.

Viechtbauer, W. 2010. Conducting Meta-Analyses in R with the Metafor Package. *Journal of Statistical Software*, 36:1–48.

Weiss, B. and M. Wagner (2011). The Identification and Prevention of Publication Bias in the Social Sciences and Economics. Journal of Economics and Statistics 231(5/6): 661-684.