Interviewer Training – Benefits and Methods

A Meta-Analysis

Jessica Daikeler  02-26-2019

Interviewer Workshop, University of Nebraska-Lincoln
• Strong link between interviewer qualification and data quality (Billiet, 1988; Dahlhamer, 2010; Olson, 2007)

• Interviewer training is an often overlooked factor in minimizing interviewer effects in interviewer-administered surveys (West and Blom, 2017).

• Huge survey projects as PIAAC (OECD, 2014) or the ESS (Loosveldt et al., 2014) as well as small projects expect well trained interviewers and survey institutes provide “trained” interviewers

• (Focus: general interviewer training, that is, the basic, cross-project part of interviewer training)
Although interviewer training is integral part of the survey process, the available literature is quite sparse.

Some research investigating the effect of interviewer training on specific data quality aspects such as unit nonresponse and correct probing (e.g., Fowler and Mangione 1990; Durand et al. 2006).

Suggestions and guidelines for interviewer training (e.g., Alcser et al. 2016; Daikeler et al. 2017).

Only Lessler, Eyerman, and Wang (2008) have provided a comprehensive qualitative overview of the literature on interviewer training.

Two focuses identifiable: Refusal Avoidance Training and data quality during the interview.
The aim of this study is to **quantify the benefits** of interviewer training and, more importantly, to **determine what aspects** of training (e.g., training length, use of blended learning, practice and feedback sessions) contribute to the **reduction of interviewer effects**.
Q1. Does general interviewer training that includes refusal avoidance training improve survey response rates compared with general interviewer training that does not include refusal avoidance training or with no interviewer training?

- Groves and McGonagle (2001, pp. 250–251) assert that two interviewer strategies—tailoring behavior to the perceived features of the sample person and maintaining interaction with the sample person—play a crucial role in gaining the cooperation of potential respondents.
- The longer the interaction lasts, the harder it is for the sample unit to refuse to participate (ebd.).

Q2. Are interviewer effects in the question-and-answer process less pronounced if the interviewers undergo training beforehand?

- Reasons for interviewer effects include the activation of social norms by the interviewer’s presence (Anderson et al. 1988; Kane and Macaulay 1993) and systematic errors in administering the survey (e.g., failure to read questions as worded, directive probing, or failure to probe; Fowler Jr. 1991, pp. 265–266).
- Interviewer training alerts interviewers to the various causes of interviewer effects with the aim of preventing, or minimizing, them.
Q3: What is the optimal interviewer training duration to reduce (a) unit nonresponse and (b) the other error sources that affect data quality?

- Learning plateau, occurs during the learning of complex skills (Thorndike 1913, p. 99)

Q4: Are unit nonresponse and interviewers’ survey administration skills in the Q&A process improved by (a) practice and feedback sessions (vs. no practice and feedback sessions); (b) interviewer monitoring (vs. no interviewer monitoring); (c) supplementary written training material (vs. no supplementary training material); (d) listening to audio refusals (vs. not listening to audio refusals); (e) blended learning (vs. an unimodal approach), and (f) previous interviewing experience (vs. no previous interviewing experience)?

- Adults learn differently than children as they accumulate their experience (Knowles 1973, p. 45)
- Most effective way of learning experiential techniques which tap the experience of the learners (visual, auditory, kinesthetic learners)
- Adults prefer self-directed, problem-centered and flexible learning
4. Literature Search Strategy

- Google Scholar, Ebsco, Web of Science, Primo, Springerlink, IPL, BL
- “Interviewer Training” OR “refusal avoidance training” OR “Refusal Aversion Training” OR (“rater training”)
- Sage Conference Abstracts, AAPOR, ESRA, JSM, WebSM, Snowballing

66 studies nested in 19 manuscripts
3. Eligibility Criteria

- Experimental Design: Treatment vs. Control or Pre/Post-Design
- Control group received no/downgraded training
- Data quality indicators need to be reported
- Survey Quality is part of interviewer training
- Refusal Avoidance training
Records identified through database searching (n = 5,527)

Additional records identified through other sources (n = 513)

Records after duplicates removed (n = 2,735)

Records screened (n = 2,735)

Records excluded (n = 2,687)

Full-text articles assessed for eligibility (n = 48)

Full-text articles excluded, with reasons (n = 29)

Full-text articles included in qualitative synthesis (n = 48)

Studies included in quantitative synthesis (meta-analysis) (66 studies nested in 19 manuscripts)
Data Generation Model
Random Effects by Hedges and Olkin (1985)
→ inference goal: generalizing beyond the studies included

Effect Size (Dependent variable) and Metric
Data Quality Percentage Difference between Trained and Untrained Interviewers
rd = Rate of Trained Interviewer – Rate of Untrained Interviewer
Effect Sizes:
Percentage of questions
- Probed correctly
- Read correctly
- Administered correctly
- Recorded correctly
- With item nonresponse

**Fig. 1** Total survey error components based on Groves and Lyberg (2010)
### 5. Examples for Effect Sizes

<table>
<thead>
<tr>
<th>Effect Size</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit nonresponse</td>
<td>Experimental interviewer group received refusal avoidance training (RAT), control group did not; number of invited vs. participating respondents in each group</td>
</tr>
<tr>
<td>Item nonresponse</td>
<td>Experimental interviewer group received advanced interviewer training, control group did not; item nonresponse rate in each group</td>
</tr>
<tr>
<td>Administering</td>
<td>Experimental interviewer group received advanced interviewer training; control group did not; number of correctly administered items per interview (audiotape error index)</td>
</tr>
<tr>
<td>Probing</td>
<td>Experimental interviewer group received advanced interviewer training, control group did not; number of correctly probed responses per interview (audiotape)</td>
</tr>
<tr>
<td>Reading out</td>
<td>Experimental interviewer group received advanced interviewer training, control group did not; number of questions correctly read out per interview (audiotape)</td>
</tr>
<tr>
<td>Recording</td>
<td>Experimental interviewer group received advanced interviewer training; control group did not; number of correctly recorded responses per interview (audiotape)</td>
</tr>
</tbody>
</table>
6. Results: Impact of Interviewer Training on Unit Nonresponse

Special RAT training improves the response rate with 7%-points.
6. Results: Factors Influencing Interviewer Training – Unit-Nonresponse

Practice and Feedback Sessions have a significant positive impact.

Practice and Feedback

P&F Session

- used listen to audio refusals
  - 0.13 [0.04, 0.21]

- Suppl. material included
  - 0.10 [0.00, 0.20]

- Monitoring included
  - 0.05 [-0.06, 0.15]

- 5 or less hours
  - 0.04 [-0.09, 0.17]

- 5 to 10 hours
  - 0.07 [-0.08, 0.22]

- More than 10 hours
  - -0.06 [-0.21, 0.10]

Interviewer training of medium-lengths is most successful for unit non-response.
6. Results: Summary for Unit-Nonresponse

**H1. How much on average do trained and untrained interviewers distinguish in unit-nonresponse rate?**
- RAT improves the unit nonresponse rate on average with 7%-points

**H2. Does it play a role for unit nonresponse of what kind of training interviewers take part? Is this finding homogenous?**
- The effect size is heterogeneous -> training characteristics do matter

**H3. What determinants render a survey unit nonresponse training successful?**
- Practice and Feedback Sessions
- Audio refusals & suppl. material
- 5-10 training hours
Training improves the item nonresponse rate with 4%-points.
6. Results: Factors Influencing Interviewer Training – Item Nonresponse

Using supplementary training material to understand the theory behind improves item nonresponse

Interviewer training of 11 hours and more is effective to gain less item nonresponse

<table>
<thead>
<tr>
<th>Training Method</th>
<th>Coefficient</th>
<th>Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training onsite only</td>
<td>-0.03</td>
<td>[-0.11, 0.04]</td>
</tr>
<tr>
<td>P &amp; F training</td>
<td>-0.04</td>
<td>[-0.12, 0.03]</td>
</tr>
<tr>
<td>Used Monitoring</td>
<td>-0.04</td>
<td>[-0.09, 0.01]</td>
</tr>
<tr>
<td>Suppl. material</td>
<td>-0.05</td>
<td>[-0.07, -0.02]</td>
</tr>
<tr>
<td>0 to 10 hours training</td>
<td>-0.00</td>
<td>[-0.04, 0.04]</td>
</tr>
<tr>
<td>11 and more training hours</td>
<td>-0.05</td>
<td>[-0.10, -0.00]</td>
</tr>
</tbody>
</table>
6. Results: Summary for item nonresponse

H1. How much on average do trained and untrained interviewers distinguish in item nonresponse rate?
- Interviewer training improves the item nonresponse rate on average with 4%-points

H2. Does is play a role for item nonresponse of what kind of training interviewers take part? Is this finding homogenous?
- The effect size is heterogeneous -> training characteristics do matter

H3. What determinants render a survey item nonresponse training successful?
- Using supplementary material
- Having longer trainings of 11 and more hours
### 6. Results: Summary

+ = tested & significant ($p \leq .005$); x = tested & not significant but expected direction; - = tested & not significant and not in expected direction; o = not tested - no variation

<table>
<thead>
<tr>
<th>Moderator/Indicator</th>
<th>Training Duration</th>
<th>Practice &amp; Feedback</th>
<th>Monitoring</th>
<th>Suppl. Material</th>
<th>Blended Learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unit-Nonresponse</td>
<td>+</td>
<td>+</td>
<td>x</td>
<td>x</td>
<td>0</td>
</tr>
<tr>
<td>Item-Administration</td>
<td>0</td>
<td>x</td>
<td>x</td>
<td>0</td>
<td>+</td>
</tr>
<tr>
<td>Item-Nonresponse</td>
<td>+</td>
<td>x</td>
<td>x</td>
<td>+</td>
<td>x</td>
</tr>
<tr>
<td>Probing</td>
<td>x</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>Recording</td>
<td>x</td>
<td>-</td>
<td>0</td>
<td>-</td>
<td>0</td>
</tr>
</tbody>
</table>
Take Home Messages

• Advanced Training improves data quality from 4 to 30% - points
• Training blocks for anti refusal training should last 5 to 10 hours while data quality training should last 11 hours and more
• Not one specific training feature that affected all data quality indicators
• Different training features, for example, practice and feedback sessions and blended learning approaches, significantly improved data quality
• Not only strongly application-oriented learning content, such as practice and feedback sessions (Knowles 1973), but also a diverse training strategy consisting of interviewer monitoring, blended learning, supplementary materials, and audio examples, are most effective.

Limitations

• Heterogeneous effect sizes problem → leads to 6 different meta-analyses with limited number of studies → low statistical power → BUT all results point in the same direction!!!
• Scope: Other data quality indicators also relevant
• Lack of variation in moderators and no experimental variation
Implications and Questions

• Interviewer training and monitoring is often **outsourced** to field institutes and is therefore difficult to influence, how can we influence interviewer training nevertheless? Any experiences?

• The use of training methods based on blended learning opens up new possibilities to create professionally developed training materials at lower costs. Any experiences with **open-access training material**?

• Further potential for better data quality undoubtedly lies in (mobile) interviewer monitoring and dashboard systems with the option of (re)training specific skills. Does anyone have experience with **targeted re-training** based on dashboard information? Does that work?

*Interested in interviewer training at Gesis? Daniela Ackermann-Piek and me are looking forward for exchange.*


Thank you for your attention.
### Backup: 5. Tasks addressed in trainings

<table>
<thead>
<tr>
<th>Interviewer task</th>
<th>Survey error potentially introduced</th>
<th>Aspects addressed in interviewer training experiments</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generate sampling frame</td>
<td>Coverage error</td>
<td>Not addressed</td>
<td>None</td>
</tr>
<tr>
<td>Make contact, gain cooperation, gain consent to additional parts of the survey</td>
<td>Unit non-response error</td>
<td>Response rate</td>
<td>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</td>
</tr>
<tr>
<td>Ask survey questions, conduct measurements and, maintain motivation</td>
<td>Measurement error and item non-response error</td>
<td>Correctly administered, read out, and probed items, item-response</td>
<td>Guest 1954; Benson and Powell 2015; Dahlhamer et al. 2010; Billiet and Loosveldt 1988; Fowler Jr. and Mangione 1986</td>
</tr>
<tr>
<td>Record answers and measurements</td>
<td>Processing error</td>
<td>Correctly recorded responses</td>
<td>Fowler Jr. and Mangione 1986</td>
</tr>
</tbody>
</table>
Training improves correct question reading/probing and answer recording with 7 - 29% points.