

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

2019 Workshop: Interviewers and Their Effects  
from a Total Survey Error Perspective

Sociology, Department of

---

2-26-2019

## Developing a Quality Control Protocol for Evaluation of Recorded Interviews

Margaret Hudson

*Survey Research Center, University of Michigan, mlhudson@umich.edu*

Lisa Holland

*Survey Research Center, University of Michigan, lisafin@umich.edu*

Lisa Lewandowski-Romps

*Survey Research Center, University of Michigan, lisalew@umich.edu*

Follow this and additional works at: <http://digitalcommons.unl.edu/sociw>



Part of the [Quantitative, Qualitative, Comparative, and Historical Methodologies Commons](#)

---

Hudson, Margaret; Holland, Lisa; and Lewandowski-Romps, Lisa, "Developing a Quality Control Protocol for Evaluation of Recorded Interviews" (2019). *2019 Workshop: Interviewers and Their Effects from a Total Survey Error Perspective*. 37.

<http://digitalcommons.unl.edu/sociw/37>

This Article is brought to you for free and open access by the Sociology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in 2019 Workshop: Interviewers and Their Effects from a Total Survey Error Perspective by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.





## Evaluating Interviewer Performance

Objectives of evaluation include:

- To measure adherence to standardized interviewing
- To provide constructive feedback to interviewers about their performance
- To create quality indicators that can be compared over time and across projects
- To take into account the needs of individual projects
- Principles of quantitative data collection should be applied to the collection of evaluation data

## Evaluation Framework

*Total Survey Error (TSE) framework*

- Measurement Error vs Errors of Representation
- Interviewer as a source of measurement error
- Standardization is primary tool for minimizing error
- Measure adherence to rules of standardization
- Intervention intended to improve performance

*Fitness for Use framework*

- Utility of data and how data are used
- Costs/resources required for implementation

## Evaluation Considerations

- Selection of questions for evaluation
- Measurement of interviewer behaviors
- Analysis of interviewer evaluation data

# Quality Control Implementation for Recorded Interviews

### Best Practices: Selection Protocols

- Select at least 5% of each interviewer's completed instruments
- Select at least some cases at random
- Select 1-2 initial interviews taken on a project for each interviewer
- Manual flags – Allow purposeful selection of cases to address concerns
- Paradata integration – Selection of cases may be informed by other interviewer performance indicators captured from ADT files or other sources (e.g., set thresholds for parameters like length of interview, short question reading time, missing data rates, etc.)
- Evaluation outcomes – For “unsatisfactory” or “needs improvement” evaluations, an additional case should be selected for evaluation.

### Measurement of Interviewer Behaviors

- Evaluation should measure interviewer adherence to standardized interviewing and any study-specific rules.
- Measurement should occur at the question level.
- Measurement should occur at the session level.
- Measures should be objective and clearly defined.
- Variation in measures should be minimized across studies.

### Question level measurement

- Question reading
- Probing for a codable answer
- Feedback to the respondent
- Entry of response

“Major” or “minor” error?

### Challenges

- Inconsistent interpretation of the major/minor distinction
- Inconsistent application of the minor error codes

How important is it to record the minor errors?

Does it matter if the interviewer makes a minor error?

### Analysis of Interviewer Evaluation Data

**Analysis showed:**

- Minor errors not driving the total error scores
- Consistent pattern across the error types (e.g., question reading, probing, data entry), with no variation in feedback errors

## Analysis (con't)

- Evaluation question lists didn't change and initial analyses used frequency of errors, allowing comparisons *within* projects only.
- However, to make comparisons between projects or even within projects if we decide to change our evaluation lists, we switched to **error ratios**.
- We use error ratios for each error type (question reading, probing, data entry, etc.).
- We also calculate an aggregate score across error types.

## Operational Implications

- Moved exclusively to use of error ratios over frequencies
- Decision to drop minor errors from scoring
- Simplifies the evaluation task
- Improves the reliability across the evaluators
- Still allows evaluators to provide feedback on occurrence of minor errors in order to correct the behavior.
- Moved to reporting by error type, not in aggregate, in order to be more purposeful about retraining or feedback to the interviewer.

## Next Steps

We will use these data to investigate the impact of our retraining protocols, to assess their effectiveness in reducing errors related to interviewer behavior.

Thanks to: Lisa Holland & Lisa Lewandowski-Romps

Contact:

**Margaret L. Hudson**  
mlhudson@umich.edu