

Does survey duration affect data quality? Evidence from survey responses

Ashiber Tsegaye Gelaw
Haleluya Tesfaye Gebru
Alemayehu Seyoum Taffesse

EEA Conference Hall | Addis Ababa, Ethiopia | July 17-18, 2026

The 23rd International Conference on the Ethiopian Economy.
Co-organized by the Ethiopian Economics Association and IFPRI-Ethiopia.



Outline

1. Motivation
2. Data and Empirical Strategy
3. Results
4. Conclusion and Policy Implications
5. Areas for Future Research



Motivation: Why Does Survey Duration Matter?

- ▶ **High-quality survey data** are essential for credible research and evidence-based policy.
- ▶ **Duration** involves a trade-off. **Existing evidence is also mixed.**
- ▶ **Large body of literature documented fatigue effects:**
 - ▶ **Abay et al. (2022):** 15-minute delay in module location ↓ dietary diversity (8–17%) and animal-source food reporting (40%).
 - ▶ **Jeong et al. (2023):** Each additional hour ↑ question skipping (up to 64%) and ↓ reported food expenditure (25%).
 - ▶ **Ambler et al. (2021):** Later interview order ↓ reported productive activities (2.2%; aggregate loss: 7.9%).

Research questions

How does interview duration affect module-level data quality, and when does additional interview time become detrimental?

Research Methodology

Data Source

- ▶ **2015 Feed the Future (FTF) Midline Survey – Ethiopia**
- ▶ Covers 5 regions, 43 zones, 84 woredas, and 250 villages
- ▶ Sample: 3,126 households and 53,149 module observations
- ▶ Retained interviews completed within a single calendar day

Research Design

- ▶ **Unit of analysis:** Module-level observation
- ▶ **Research approach:** Quantitative research approach



Measurement of Key Variables

Survey Duration

- ▶ Measured as module completion time (minutes) using CAPI timestamps.

Data Quality

- ▶ Constructed from six indicators:
 - ▶ **Completeness**
 - ▶ **Consistency**
 - ▶ **Skip-pattern**
 - ▶ **Outliers**
 - ▶ **Validity**
 - ▶ **Straight-lining**
- ▶ Two measures:
 - ▶ **Rule-based index (main):** Ordered quality categories (Very poor to Very good).
 - ▶ **PCA index (robustness):** Continuous composite quality score.



Econometric Model Specification

- ▶ Based on respondent burden and satisficing theories, we specify the following Mixed-effects ordinal logistic model:

$$\begin{aligned} DQ_{imj}^* = & \beta_0 + \beta_1 \ln(Dur_{imj}) + \beta_2 [\ln(Dur_{imj})]^2 \\ & + \beta_3 Order_{mj} + \beta_4 (\ln(Dur_{imj}) \times Order_{mj}) \\ & + \beta_5 X_{ij} + \alpha_m + V_j + \varepsilon_{imj} \end{aligned} \quad (1)$$

- ▶ The latent quality measure is linked to five observed categories:

$$DQ_{imj} = \begin{cases} 1 & \text{Very Poor: } DQ_{imj}^* \leq \mu_1 \\ 2 & \text{Poor: } \mu_1 < DQ_{imj}^* \leq \mu_2 \\ 3 & \text{Fair: } \mu_2 < DQ_{imj}^* \leq \mu_3 \\ 4 & \text{Good: } \mu_3 < DQ_{imj}^* \leq \mu_4 \\ 5 & \text{Very Good: } DQ_{imj}^* > \mu_4 \end{cases}$$

Research Hypotheses

- ▶ **H1 (Nonlinear duration effect):** Interview duration has an inverted U-shaped relationship with data quality.
- ▶ **H2 (Survey fatigue):** Later interview modules exhibit lower data quality due to respondent fatigue.
- ▶ **H3 (Duration $\times$ Order):** The adverse effect of longer interview duration is stronger for modules administered later in the interview.

Robustness Checks

- ▶ Generalized ordered logit and mixed-effects Probit models.
- ▶ Alternative duration specifications and trimming of duration outliers.
- ▶ Multicollinearity assessment among explanatory variables.
- ▶ IV control function approach using average enumerator interview time.
- ▶ Jackknife re-estimation excluding enumerators sequentially.



Descriptive Statistics

Variable	Statistic	Value
Sample	Module-level observations	53,149
Households		3,126
Enumerators		47
Respondent age	Mean (SD)	44.9 (14.9)
Female respondents	Share	30.7%
Married (monogamous)	Share	70.7%
Module duration (minutes)	Mean (SD)	17.6 (28.3)
	Range	1–298
Questions per module	Mean (SD)	92.0 (156.0)
Data quality	Very poor–Very good	0.7–0.2%
	Poor	13.3%
	Moderate	61.3%
	Good	24.7%

Ordered Logistic Regression Models: Data Quality

	M1	M2	M3	M4	M5	M6	M7
$\ln(\text{Duration})$	-0.506***	-0.472***	-0.479***	1.035***	1.061***	0.379***	0.961***
$\ln(\text{Duration})^2$	—	—	—	-0.353***	-0.491***	-0.032*	-0.173***
Middle module	—	0.038	0.032	-0.214***	-2.783***	—	-10.687***
Late module	—	0.235***	0.230***	0.310***	-0.358***	—	-0.225***
Middle $\times$ Duration	—	—	—	—	1.227***	—	0.261***
Late $\times$ Duration	—	—	—	—	0.411***	—	-0.894***
Respondent controls	No	No	Yes	Yes	Yes	Yes	Yes
Module fixed effects	No	No	No	No	No	Yes	Yes
Enumerator random effects	No	No	No	No	No	Yes	Yes
Observations	53,149	53,149	53,149	53,149	53,149	53,149	53,149

Preferred Model

Model 7.



Optimal Survey Duration

- ▶ The nonlinear model identifies an **optimal duration threshold**, beyond which additional interview time reduces data quality.
- ▶ **Optimal duration by module order:**

Module Position	Minutes
Early	2.95
Middle	10.30
Late	4.50

Key Implication

Optimal survey length is context dependent.

Conclusion

- ▶ **Duration matters:** Interview time has a nonlinear effect—moderate duration improves quality, while excessive length reduces quality.
- ▶ **Context matters:** Module position, complexity, and questionnaire design shape fatigue effects.
- ▶ **Interviewers matter:** Enumerator practices contribute significantly to data quality differences.
- ▶ **No universal threshold:** Optimal duration depends on module characteristics and respondent burden.
- ▶ **Robust evidence:** Findings remain stable across alternative models and sensitivity checks.

Policy Implications

- ▶ **Survey design:** Balancing information needs and respondent burden is central to improving data quality.
- ▶ **Questionnaire development:** Module sequencing and survey complexity influence measurement accuracy.
- ▶ **Field implementation:** Enumerator capacity and quality monitoring remain critical for reliable data collection.

Key Message

Better survey quality comes from managing burden, not simply shortening interviews.

Areas for Future Research

- ▶ **Generalizability:** Validate findings across countries, populations, and survey modes.
- ▶ **Measurement:** Integrate external validation data and improved quality metrics.
- ▶ **Mechanisms:** Investigate how cognitive burden, motivation, topic sensitivity, and interviewer behavior drive fatigue effects.
- ▶ **Causal evidence:** Use experimental and panel designs to identify optimal duration thresholds and alternative burden measures (e.g., questions, pages, and content complexity).

Thank You

