Reading Fast, Reading Slow:
Interviewer Behavior and the Influence on Survey Outcomes

Michael Bergmann
Technical University of Munich (Chair for the Economics of Aging),
Munich Center for the Economics of Aging (MEA)
Max Planck Institute for Social Law and Social Policy

Johanna Bristle*
Munich Center for the Economics of Aging (MEA),
Max Planck Institute for Social Law and Social Policy

27th International Workshop on Household Survey Nonresponse,
31.08-2.9.2016, Oslo, Norway

1. Introduction

Standardized interviewing aims at keeping the interview situation for all respondents constant in order to minimize interviewer influence (Groves, et al., 2009). One key component of standardized interviewing in face-to-face surveys is the reading behavior of the interviewer, especially the pace by which interviewers read out questionnaire items to respondents. Studies showed that interviewers do not follow the rules of standardized interviewing, but instead change their interviewing behavior as they gain more experience over the survey's field period (Cannell, et al., 1977, Fowler, 1991, Olson and Peytchev, 2007).

The central aim of this paper is to shed light on the question, if and how a change in interviewers' reading times and thus a deviation from standardized interviewing over time matter with respect to substantive findings. Especially in a cross-national setting, comparability and standardization of interviewing are of utmost importance. We formulate three hypotheses that we will test in our study.

1. We expect that only a small proportion of interviewers will read question texts exactly as they are worded. Instead, interviewers will adjust their reading behavior over the survey's field period resulting in a decrease of reading time (shortening effect).
2. Following an argument that it is a rational behavior to shorten interviews given a payment structure by interview instead of by hour, we expect that such a decrease of reading time is uniform across countries and thus can be generalized.
3. We further hypothesize that, depending on the amount of informational content, which is shortened or read out very fast, this will affect measurement and thus substantive findings to a varying extent. A strong decline in reading time as well as a very short reading time should then be associated with less reliable and less informed responses.

*Corresponding author: Johanna Bristle; e-mail: bristle@mea.mpisoc.mpg.de.

2. Data

We use data from wave 5 Release 1.0.0 of the Survey of Health, Ageing and Retirement in Europe (SHARE; Börsch-Supan, et al., 2013, Börsch-Supan, 2015). SHARE is a cross-national, ex-ante harmonized panel survey In addition, latent timers were extracted from Blaise audit trails. They measure the exact duration from when the question appears on the screen until the answer is entered by the interviewer. Our final analysis sample consists of 62536 interviews that have been conducted by 1582 interviewers in 15 countries.

Our main variable of interest is the reading duration of specific items. To obtain a clean measure of interviewer behavior, we extracted only those items that contained mere introduction or explanation texts to be read by the interviewer and hence barely any interaction between interviewer and respondent. We then selected six items, which were not backed up, edited, or re-entered by the interviewer, and that contain relevant information for the respondents to provide a well-informed answer. Table 1 gives a short overview on the used introductions and the corresponding survey outcomes.

Table 1: List of used introduction variables and the respective survey outcomes

Type of content	Reading item	Survey outcome
Confidentiality	Intro to overall SHARE interview	Nonresponse to income question
Confidentiality	Intro to record linkage	Consent given to record linkage
Definition	Intro to health care expenditures	Payed anything out-of-pocket
Definition	Intro local area information	Feeling part
		Cleanliness
		Help available
		Vandalism or crime
Instruction	Intro to recall test	Amount of words recalled
Instruction	Intro to chair stand	Compliance with chair stand test

3. Empirical Model

In our data, we observe interviews over an interviewer's field period. Thus, the data collected by the interviewers can be written in an unbalanced panel structure. Throughout our analyses, we use fixed effects regression models (Allison, 2009, Wooldridge, 2013) to control for time invariant differences between interviewers and to focus on intra-individual changes of interviewers' reading behaviors over the survey's field period.

In general, we expect the reading behavior to change over the field period. We expect a sharp decrease in interviewers' reading time during the first interviews due to increasing experience and a more stable pattern later on. Therefore, we model the number of interviews in a semiparametric way by using piecewise regression lines, called linear splines (for an introduction, see Keele, 2008). We set spline knots at the 2nd, the 10th, and the 50th interview. From a substantial point of view, the most interesting time interval is between the 2nd and the 10th interview. In this period we expect interviewers to adjust their reading behavior the most.

The fixed effects approach enables us to disentangle the mechanisms potentially leading to a decline in reading time as the fixed effects transformation allows us to neglect all stable interviewer characteristics (time-invariant). After carefully controlling for potential

time-variant confounders (period effects, respondent characteristics, interview situation) and explicitly estimating change in reading behavior using splines, we are confident to ascribe a negative effect of the increasing number of conducted interviews, especially between the 2nd and 10th interview, to an interviewer's learning behavior to shorten, skip or speed through the item.

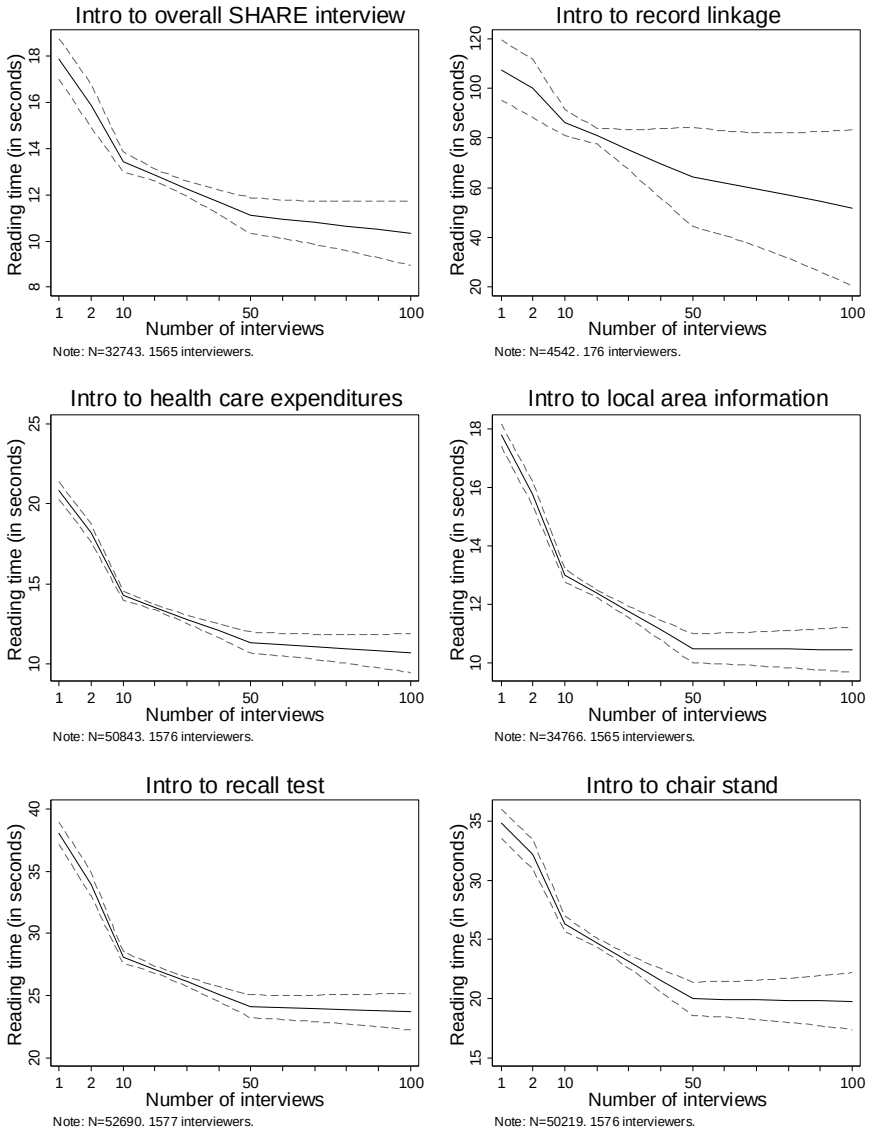
In the second step of our empirical analyses, our main variable of interest, i.e. the changing interviewer's reading duration, acts as independent variable in order to explain substantial consequences on the survey outcomes displayed in Table 1 that then serve as dependent variables. Otherwise, our general model stays the same. Thus, we again run linear fixed effects regression models and control for period effects, confounding effects of the changing sample composition including respondent characteristics, and aspects of the specific interview situation that might differ between interviews.

4. Results

4.1. Change of interviewers' reading behavior over the survey's field period

In Figure 1, predicted reading time is displayed over the number of completed interviews with 95% confidence intervals. To highlight the spline structure, the x-axis is stretched in the beginning. Thus, the decline in this interval is even steeper than it appears in the graph. Overall, we see a strong decline for all six introduction items of SHARE. The steepest decrease can be observed in the beginning. A change within the 2nd and the 10th interview is what we term a learning effect of the interviewer, which is negative and significant for all selected items. To give a concrete example: The introduction to the overall SHARE interview has a mean reading time of about 18 seconds at the first interview and drops to about 13 seconds at the tenth interview, which is a decrease in reading time of almost 30 percent. Until the 50^{th} interview, the decrease is still significant but attenuated a bit. Between the 50^{th} and the 100^{th} interview all six observed items only show marginal decreases, while at the same time standard errors increase due to the lower number of interviewers having such a large workload.

Figure 1: Change in reading behavior over number of interviews

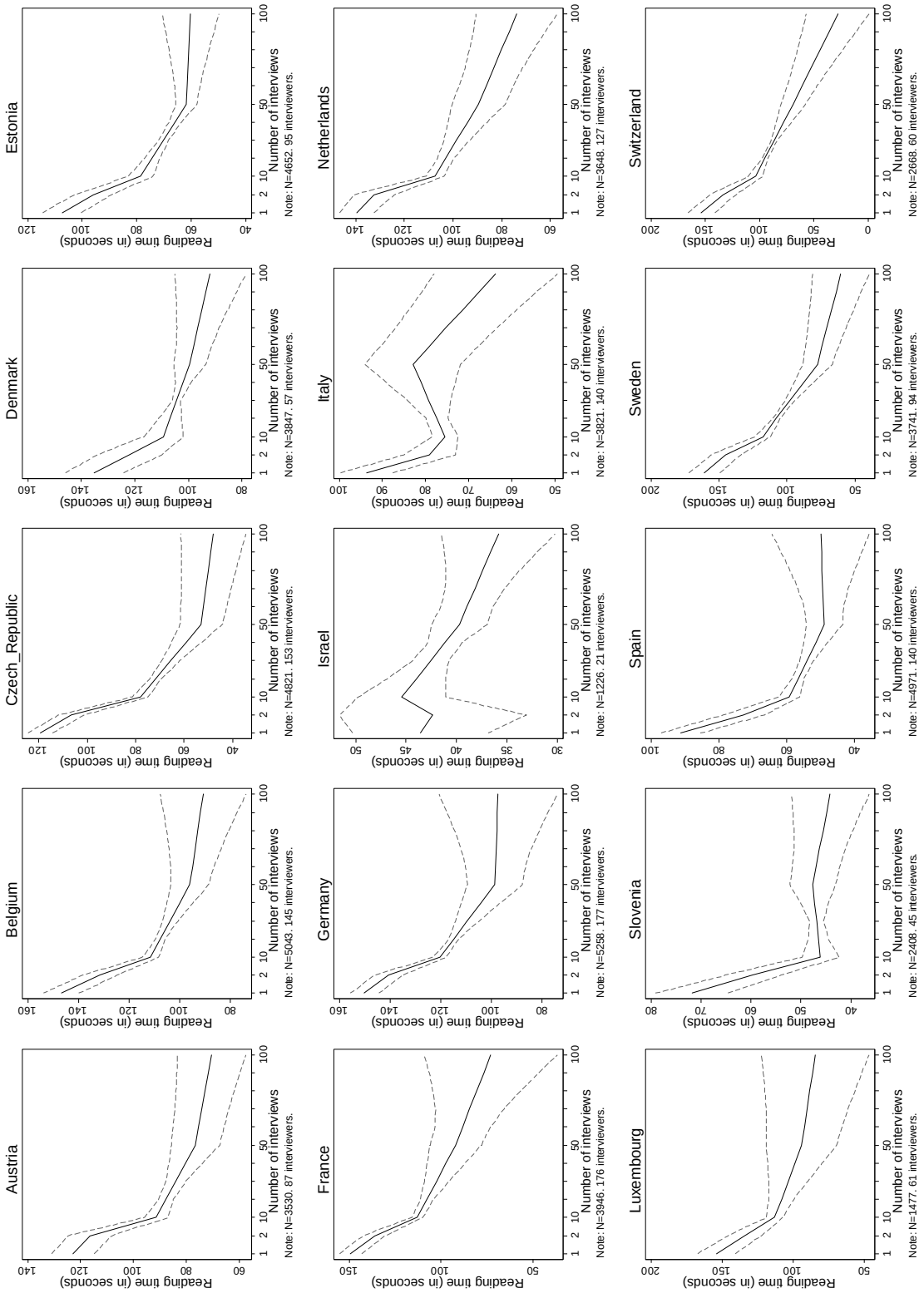


Data: SHARE wave 5.Fixed effects regressions with spline knots at 2, 10, and 50 interviews.
Controlled for days in field, respondent characteristics, and interview specifics. Weighted results.

4.2. Uniform decline across countries

When we allow for country-specific heterogeneity, the overall picture stays more or less the same. Figure 2 shows the cumulated reading time of all 6 intro texts. Overall, we see a comparable strong decrease in interviewers' reading time within the first ten interviews. What attracts attention is the very different (starting) level of interviewers' reading time that can be attributed mostly to language differences.

Figure 2: Country-specific change in reading behavior over number of interviews



Data: SHARE wave 5.Fixed effects regressions with spline knots at 2, 10, and 50 interviews.
Controlled for days in field, respondent characteristics, and interview specifics. Weighted results.

4.3. Does it matter? Consequences for survey outcomes

We further analyzed to what extent the observed decline in reading time substantially affects survey outcomes. Therefore, we use the reading time of the selected introduction texts as main explanatory variables, while our dependent variables are the survey outcomes described in Table 1 above. We again run fixed effects models and control for period effects, respondent characteristics, and the interview situation the same way as before. The results are presented in Table 2. The coefficients presented in the last column equal the effect of the average change in reading time in the sample and can be interpreted as change in percentage points, i.e. the average decrease in reading time with respect to the record linkage item reduces the consent rate by five percentage points for example. When looking at country-specific models, the main results are resembled for nearly all survey outcomes under consideration.

Table 2: Intro-specific regressions on survey outcomes

Reading item	Survey outcome	Avg. change in reading time (in seconds)	Effect of avg. change
Intro to overall SHARE interview	Refusal to income [0;1]	-5.4	-0.000 (0.001)
Intro to record linkage	Consent given [0;1]	-55.1	-0.050*** (0.008)
Intro to health care expenditures	Payed anything out of pocket [0;1]	-12.2	-0.006 (0.003)
Intro to local area information	Feeling part [0;1]	-8.3	0.000 (0.002)
	Cleanliness [0;1]		0.005 (0.003)
	Help available [0;1]		-0.005 (0.003)
	Vandalism or crime [0;1]		0.008* (0.003)
Intro to recall test	Amount of words [0;10]	-18.4	-0.018* (0.008)
Intro to chair stand	Compliance with test [0;1]	-16.9	-0.024*** (0.001)

Note: Each line represents an own linear fixed effects model on the respective survey outcome with reading time of intro text as explanatory variable and days in field, respondents characteristics, and interview specifics as controls.
Panel-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

5. Discussion

Our analyses show that, in contrast to the goal of standardized interviewing, interviewers' reading time significantly decreases over the survey's field period. Of course, a decrease in reading time alone must not be problematic for data quality per se. Therefore, the most important question we tried to answer in this article is: does it matter? Our findings reveal a rather large effect of decreasing reading time on within-survey requests, such as the consent given by the respondents to link their survey answers to administrative data or doing a physical test to measure the respondents' endurance. Reading times matter less for skipping other survey intros that are followed by subjective evaluations.

References

Allison, Paul D. 2009. *Fixed Effects Regression Models*. Thousand Oaks, CA: SAGE Publications.

Börsch-Supan, Axel. 2015. "Survey of Health, Ageing and Retirement in Europe (SHARE) Wave 5. Release Version: 1.0.0." SHARE-Eric. Data Set.

Börsch-Supan, Axel, Martina Brandt, Christian Hunkler, Thorsten Kneip, Julie Korbmacher, Frederic Malter, Barbara Schaan, Stephanie Stuck, and Sabrina Zuber. 2013. "Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE)." *International Journal of Epidemiology* 42: 992-1001.

Bradburn, Norman M., and Seymour Sudman. 1979. *Improving Interview Method and Questionnaire Design: Response Effects to Threatening Questions in Survey Research*. San Francisco, CA: Jossey-Bass.

Cannell, Charles F., Kent H. Marquis, and André Laurent. 1977. "A Summary of Studies of Interviewing Methodology." Washington, DC: US Government Printing Office.

Chromy, James R., Joe Eyerman, Dawn Odom, Madeline E. McNeeley, and Art Hughes. 2005. "Association between Interviewer Experience and Substance Use Prevalence Rates in Nsduh." In *Evaluating and Improving Methods Used in the National Survey on Drug Use and Health*, edited by Joel Kennet and Josph Gfroerer, 59-87. Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.

Fowler, Floyd J. 1991. "Reducing Interviewer-Related Error through Interviewer Training, Supervision, and Other Means." In *Measurement Errors in Surveys*, edited by Paul P. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz and Seymour Sudman, 259-78. Hoboken, NJ: John Wiley & Sons.

Groves, Robert M., Floyd J. Fowler, Mick P. Couper, James M. Lepkowski, Eleanor Singer, and Roger Tourangeau. 2009. *Survey Methodology*. 2 ed. Hoboken, NJ: John Wiley & Sons.

Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics* 3: 305-60.

Keele, Luke John. 2008. *Semiparametric Regression for the Social Sciences*. Hoboken, NJ: John Wiley & Sons.

Kosyakova, Yuliya, Jan Skopek, and Stephanie Eckman. 2015. "Do Interviewers Manipulate Responses to Filter Questions? Evidence from a Multilevel Approach." *International Journal of Public Opinion Research* 27: 417-31.

Olson, Kristen, and Andy Peytchev. 2007. "Effect of Interviewer Experience on Interview Pace and Interviewer Attitudes." *Public Opinion Quarterly* 71: 273-86.

Wooldridge, Jeffrey. 2013. *Introductory Econometrics: A Modern Approach*. 5 ed. Mason, OH: South-Western Cengage Learning.