methods, data, analyses | Vol. 11(2), 2017, pp. 163-188 DOI: 10.12758/mda.2017.03

How Interviewer Effects Differ in Real and Falsified Survey Data: Using Multilevel Analysis to Identify Interviewer Falsifications

Uta Landrock
University of Kaiserslautern

Abstract
In face-to-face interviews, interviewers can have an important positive influence on the quality of survey data, but they can also introduce interviewer effects. What is even more problematic is that interviewers may decide to falsify all or parts of interviews. The question that the present article seeks to answer is whether the interviewer effects found in falsified data are similar to those found in real data, or whether interviewer effects are larger and more diverse in falsified data and may thus be used as an indicator for data contamination by interviewer falsifications. To investigate this question, experimental data were used from controlled real interviews, interviews falsified by the same interviewers, and questionnaires completed by these interviewers themselves as respondents. Intraclass correlations and multilevel regression models were applied, and interviewer effects in the real survey data were compared with those in the falsified data. No evidence of interviewer effects was found in the real data. By contrast, interviewer effects were found in the falsified data. In particular, there was a significant association between the interviewers' own responses and the falsified responses to the same questions in the questionnaire. Thus, to detect interviewer falsifications, I recommend that researchers should also get the interviewers to complete the questionnaire and check datasets or suspicious cases for interviewer effects.

Keywords: interviewer, interviewer effects, interviewer falsifications, data quality, identification of falsifications, multilevel analysis

1 Introduction

Face-to-face interviews are an important mode of data collection in empirical social research. It is used in many major studies, for example the European Values Study (EVS),[1] the U.S. General Social Survey (GSS),[2] and the Programme for the International Assessment of Adult Competencies (PIAAC).[3] Interviewers can have a major influence on the quality of survey data. On the one hand, they can improve data quality, for example by helping the respondent to understand the survey questions correctly (Mangione, Fowler, & Louis, 1992). On the other hand, there is the risk of interviewer effects, that is, distortions of survey responses due to the presence of an interviewer. Interviewer effects can cause biased data and affect substantive findings (Beullens & Loosveldt, 2016; Groves & Magilavy, 1986). They occur when the respondent's answer depends not only on the intended stimulus of the question but also on the interview situation and the interviewer (Bogner & Landrock, 2016; Schanz, 1981). In the case of interviewer effects, certain interviewer behaviors (e.g., reading pace or suggestiveness) or characteristics (e.g., experience, age, gender, or education) may influence the response behavior of the respondent (Beullens & Loosveldt, 2016; Haunberger, 2006; Mangione et al., 1992). Interviewer effects therefore constitute response bias (see Groves & Magilavy, 1986), where the reported values of the respondent systematically deviate from the true values.

In this context, it is important to know whether some types of questions are more susceptible to interviewer effects than others (Mangione et al., 1992). Research on interviewer effects has yielded a large number of findings in this regard (for an overview, see Bogner & Landrock, 2016). According to Haunberger (2006), for example, difficult and sensitive questions, attitudinal questions, and open-ended questions are particularly prone to interviewer effects. Haunberger (2006) showed that, in the case of difficult questions, the gender and education of the interviewers may have an influence on responses, for example, to income-related questions. The probability that the respondent will refuse to answer such questions is reported to

1 http://www.europeanvaluesstudy.eu/
2 http://gss.norc.org/
3 http://www.oecd.org/skills/piaac/

Acknowledgements

The research presented in this paper was funded by a German Research Foundation (DFG) grant awarded for the project IFiS – Identification of Falsifications in Surveys (WI 2024/5-4 and ME 3538/4-1). This financial support is gratefully acknowledged. I would also like to thank the referees and the editors for their helpful comments on an earlier version of the manuscript.

Direct correspondence to

Uta Landrock, University of Kaiserslautern, Erwin-Schroedinger-Straße 57,
67653 Kaiserslautern, Germany
E-mail: landrock@sowi.uni-kl.de

be higher in the case of female or highly educated interviewers (Bogner & Landrock, 2016; Haunberger, 2006). Regarding attitudinal questions, research findings are ambiguous. Whereas Liu and Stainback (2013) identified interviewer gender effects on responses to attitudinal questions, Groves and Magilavy (1986) did not find evidence of such an influence on attitudinal questions compared to factual questions. Haunberger (2006) suggested that interviewer age and education may influence responses to open-ended questions and that these questions are therefore susceptible to interviewer effects (Mangione et al., 1992). By contrast, Groves and Magilavy (1986) reported that open-ended questions were not inherently more susceptible to interviewer effects than closed questions. However, in the case of open questions that ask respondents to mention several entities, for example "What do you think are the most important problems facing the country?," the authors suggested that the likelihood that the respondent would mention a second entity might depend on the interviewer's probing behavior, and that "the differential behaviors that determine whether a second mention is given also might influence substantive responses on the second mention" (Groves & Magilavy, 1986, p. 260). In summary, therefore, research findings show that difficult, attitudinal, and open-ended questions are susceptible to interviewer effects.

These findings provide evidence that the perceptible sociodemographic characteristics of the interviewer – namely gender, age, and education – are relevant to the occurrence of interviewer effects (Haunberger, 2006; Liu & Stainback, 2013; West & Blom, 2016). Olson and Bilgen (2011) reported that larger interviewer effects occurred with respect to acquiescence in the case of experienced interviewers than in the case of inexperienced interviewers. West and Blom (2016) described the influence of certain personality traits of the interviewers that may affect response behavior. Moreover, research findings suggest that the relation between interviewers' and respondents' characteristics may result in interviewer effects: Schanz (1981) analyzed the relevance of interaction effects and described positive correlations between the answers of the interviewer and the answers of the respondent to the same survey questions. One possible explanation for this positive correlation is that the respondent reacts to the non-verbally expressed attitudes of the interviewer (Schanz, 1981; West & Blom, 2016). Thus, interviewer effects may also depend on the content of the question and the interaction of the attitudes of the interviewers and the respondents (Schanz, 1981).

In face-to-face interviews, not only may interviewer effects occur, but interviewers may even decide to falsify all or parts of interviews. This is the most extreme and problematic form of influence that an interviewer can exert. Falsifications may severely bias the results of analyses and lead to incorrect results (Landrock, 2017; Reuband, 1990; Schnell, 1991; Schraepler & Wagner, 2003). A reliable strategy for identifying falsifications would therefore be extremely valuable to ensure high quality in interviewer-based survey research. However, research has shown that, based on univariate distributions (Menold & Kemper, 2014; Reuband, 1990; Schnell, 1991) and multivariate correlations (Landrock, 2017), falsified and real data appear to be quite similar and that the existence of falsifications in data is thus not readily noticeable. Given that the falsification of interviews may be considered to be an extreme form of interviewer effect, statistically testing for interviewer effects might provide a more effective indicator for identifying falsifications. This paper therefore analyzes and compares interviewer effects in real survey data and in data falsified by interviewers. Using experimental data, the aim is to determine whether similar interviewer effects occur in falsified data and in real data or whether interviewer effects are larger and more diverse in falsified data and may thus be used as an indicator for data contamination by interviewer falsifications (see Winker, Kruse, Menold, & Landrock, 2015).

In falsified interviews, by definition, no interaction takes place between the respondent and the interviewer. Therefore, it may seem implausible to assume that interviewer effects occur in a dataset comprised of falsified data. However, in falsified interviews, interviewers obviously have a direct influence on the data reported as answers by the respondent. Yet, they have only a little information about the respondent. Consequently, the fabrication of plausible responses depends very strongly on the falsifier. Thus, interviewer effects – or, more precisely, "falsifier effects" – can be expected.

Different falsifiers may falsify the respondents' answers in different ways. It is conceivable that certain socioeconomic, demographic, or psychological characteristics of the falsifiers may find their way into the data they falsify. Both the falsifiers' perceptions of social reality and their falsifications are influenced by personal characteristics. Therefore, the interviewers' characteristics should be significant explanatory variables in a dataset that is contaminated by interviewer falsifications. Moreover, I assume that interviewer effects are more pronounced in falsified than in real survey data (see Winker et al., 2015).

In the research presented in this paper, a number of variables that are known to be generally susceptible to interviewer effects are analyzed as dependent variables with the aim of determining (a) the degree to which interviewer effects occur in real and in falsified data and (b) whether there are differences between the interviewer effects in real and in falsified survey data.

2 Hypotheses

To contribute to research on interviewer effects, to knowledge of interviewer falsifications and their impact on data quality, and to potential strategies for identifying contaminated data, the following two general hypotheses will be tested:

H1: Interviewer effects occur both in real and in falsified data.

As falsifying interviewers have only a little information about the respondent, they must draw on their personal experience of social reality in order to fabricate plausible answers to survey questions. Thus, interviewer effects may occur not only in real survey data but also in falsified survey data (see Winker et al., 2015).

H2: The interviewer effects in falsified data are larger than in real data.

I assume that sociodemographic or psychological characteristics of interviewers are more likely to find their way into falsified survey data than into real data.

Regarding the interviewer characteristics that may cause interviewer effects or influence the way in which an interviewer falsifies, explanatory variables will be analyzed that can theoretically be expected to be susceptible to interviewer effects. The following more specific hypotheses will be tested on real data and on falsified data:

H3a: The core sociodemographic characteristics of the interviewers affect the reported responses.

As reported by West and Blom (2016), Haunberger (2006), Mangione et al. (1992), and Liu and Stainback (2013), sociodemographic characteristics of the interviewer – in particular gender, age, and education – may lead to interviewer effects. I further expect that income, as an indicator of socioeconomic background, may also cause interviewer effects.

H3b: The magnitude of interviewer effects depends on the interviewer's experience.

Olson and Bilgen (2011) found that experienced interviewers caused larger interviewer effects than inexperienced interviewers. Hypothesis H3b will test whether this finding is replicated in the present study.

H3c: Associations exist between the behaviors and attitudes of interviewers and the reported behaviors and attitudes of the respondents they interview.

Following Schanz (1981), I assume that associations will be found between the answers of the interviewers and the answers of the respondents to the same survey question – in other words, that the interviewer's response to the same survey question affects the response reported by the respondent.

H3d: The occurrence and magnitude of interviewer effects depends on the personality traits of the interviewer.

Both West and Blom (2016) and Winker et al. (2015) found evidence that suggested that the personality traits of the interviewer may lead to interviewer effects. West and Blom (2016) reported an effect of interviewers' extraversion and self-confidence. Accordingly, I assume that interviewers with higher levels of extraversion produce larger interviewer effects than introverted interviewers. By contrast, more conscientious interviewers should produce smaller interviewer effects than interviewers with a lower level of conscientiousness. With regard to self-confidence, I assume that interviewers with a higher level of perceived self-efficacy perform better, and therefore produce smaller interviewer effects, than interviewers with a lower level of perceived self-efficacy.

H3e: The magnitude of interviewer effects depends on the interviewer payment scheme used (payment per completed interview vs. payment per hour).

In their study of interviewer effects in real and falsified interviews, Winker et al. (2015) found that the payment scheme (i.e., the type of monetary compensation) applied had an impact on the collected data and therefore on the quality of a survey. I assume that interviewers who are paid per completed interview produce larger interviewer effects than interviewers paid per hour. Winker et al. (2015) also found correlations between the payment scheme and political participation (operationalized as the number of political activities mentioned by the respondent). For the real data, the authors showed that payment per hour was associated with a higher number of political activities mentioned. It would appear that payment per hour leads to more complete data and thus to higher data quality. Hypothesis H3e will test the assumption that interviewers who are paid per completed interview produce larger interviewer effects than interviewers who are paid per hour.

3 Data Base and Methods

Due to the virtual non-existence of datasets with proven falsified interviews, experimental data were used to analyze falsified data and their differences to real data (see Winker et al., 2015). My data base comprised three datasets. The data were collected at the University of Giessen, Germany in summer 2011 in the framework of the research project IFiS – Identification of Falsifications in Surveys (see also Menold & Kemper, 2014; Winker et al., 2015).

In the first step, 78 interviewers conducted 710 real face-to-face interviews. The questionnaire consisted of 62 questions, which were taken mainly from the 1998 German General Social Survey (ALLBUS) questionnaire.[4] Besides sociodemographic questions, the questionnaire comprised attitudinal and behavioral items on social, political, and economic topics. The average interview duration was 30 minutes. Both the respondents and the interviewers were students at the University of Giessen. The interviewers themselves selected the respondents on the university campus without any quota restrictions and interviewed them. The audio-recorded interviews were checked to make sure that they had been conducted correctly. Half

4 http://www.gesis.org/en/allbus/allbus-home/

of the interviewers were paid per completed interview (8 euros), the other half were paid per hour (12 euros). Prior to data collection, an interviewer training session was conducted, in the course of which the interviewers were familiarized with the research design and the questionnaire.

For the second dataset, 710 interviews were fabricated. For this purpose, the same interviewers who had conducted the real interviews were requested to fabricate survey data in the lab. Hence, for each real interview, a corresponding fabricated interview was obtained. Compensation was allocated either per interview (3 euros per falsified interview) or per hour (9 euros per hour). The falsifying interviewers were given details of the sociodemographic characteristics of the persons whose interviews they were to fabricate. These persons were real survey participants, who had been interviewed previously by another student interviewer. The information provided included the respondent's gender, age, subject studied, number of semesters enrolled, marital status, place of residence, living situation (i.e., the person or persons with whom the respondent lived in a household), and country of origin. In the case of a genuine (i.e., uninstructed) falsification in an actual fieldwork setting, the falsifying interviewer could easily have obtained this information by briefly interviewing the respondent. The falsifiers were requested to imagine the described person and to complete the questionnaire, thus fabricating the data as if they had been collected in a real survey fieldwork setting.

The exact instructions for falsifying an interview were:

> Please read carefully the description of the person whose interview you are to falsify. Please complete the attached questionnaire as if you had really conducted a personal interview with the respondent. During falsification, please place the description of the respondent next to the questionnaire, so that you are always aware of the characteristics of that person.
>
> The person whose interview you are to falsify…
>
> - is female,
> - is 20 years old,
> - studies teaching,
> - is enrolled in her second semester at a university.
> - She is unmarried, in a steady relationship,
> - lives in Huettenberg, a rural village in Hesse,
> - with her parents or relatives.
> - Country of birth: Germany.

As a last step, the interviewers themselves, as respondents, completed the same questionnaire that they had previously used for interviewing and falsifying. These self-administered interviews generated the third dataset.

This experimental setup has strengths, but it also has weaknesses. One weakness is that the respondents and interviewers were students and that core sociodemographic characteristics, such as age and education, therefore displayed only small variance (see Winker et al., 2015). The major strength of the experimental setup, compared to a standard field setting, was the possibility of collecting more information about the interviewers and their falsifying processes. Because they were surveyed with the same questionnaire as the proper respondents, the dataset includes not only information about respondents and fictitious respondents but also about the interviewers. This offers great potential for analyzing interviewer effects.

There are several possible approaches to investigating interviewer effects. Schanz (1981) analyzed the influences of interviewer characteristics on the response behavior of the participants by estimating multiple regression analyses. First, he included substantive explanatory variables; then he added interviewer variables. Mangione et al. (1992) and Groves and Magilavy (1986) measured interviewer effects by intraclass correlation. The intraclass correlation expresses the proportion of the item variance that is attributable to the interviewer (Mangione et al., 1992). In the absence of interviewer effects, the value of the intraclass correlation should be zero or close to zero (Beullens & Loosveldt, 2016). Olson and Bilgen (2011) estimated multilevel regression analyses with respondent characteristics such as age and education on the respondent level (individual level) and interviewer characteristics such as age, education, and experience on the interviewer level (contextual level).

At first glance, it would appear to be useful to estimate ordinary least squares (OLS) regressions. However, especially when it comes to analyzing interviewer effects, it makes sense to assume that – as expressed in the above-mentioned hypotheses – the observations of the respondents (i.e., the individual interviews) are probably not independent from the interviewers. Therefore, the model assumptions of OLS regressions are not met. Rather, the data are organized hierarchically, and multilevel regression analyses are thus more appropriate (Hox, 1995). The respondents represent the individual level, and the interviewers represent the group or contextual level.

To investigate the impact of interviewer characteristics on substantive findings, intraclass correlations were also estimated and multilevel regression analyses were conducted. To answer the research question as to what influence interviewers have on the data and findings and whether there are differences between real and falsified data in this respect, identical multilevel regression models were estimated separately with real and with falsified data. Thus, to determine what differences occur, the respective results – in particular, the effects of the various independent variables – were compared. This approach also allowed the identification of interviewer effects on substantive findings.

Table 1 Overview of variables used to analyze interviewer effects

Dependent Variables	Independent Variables on the Individual (Respondent) Level	Independent Variables on the Contextual (Interviewer) Level
Income	Age Living situation	Payment scheme
Political participation	Gender Internal political efficacy Political dissatisfaction Extremism	Interviewer's gender Interviewer's income Interviewer's response to the same questions of the questionnaire
Political anomy	Economic dissatisfaction External political efficacy	Interviewer's experience
Healthy eating behavior	Intention Perceived behavioral control TV consumption Body mass index Doing sports Preference for healthy desserts	Interviewer's extraversion Interviewer's conscientiousness Interviewer's level of perceived self-efficacy

4 Operationalization and Multilevel Regression Model

Table 1 gives an overview of the dependent and independent variables used. These variables are explained in more detail in the following sections.

4.1 Dependent Variables on the Individual Level

One aim of the present study was to analyze a number of dependent variables that I considered to be particularly susceptible to interviewer effects, namely (a) income, as a sensitive (and open-ended) factual question; (b) political participation, as a behavioral question; (c) political anomy, as an attitudinal question; and (d) healthy eating behavior, as an additional behavioral question.

Income was measured with the question: "How much money is at your disposal on average per month, during the current semester?"

Political participation was measured using a list of twelve political activities. The wording in the questionnaire was:

> If you wanted to have political influence or to make your point of view felt on an issue that was important to you: Which of the possibilities listed on these cards would you use? Which of them would you consider? Please name the corresponding letters.

In a previous study, I analyzed the effects of falsified data on the results of multivariate theory-driven OLS regression analyses, using the explanation of political participation as an example (Landrock, 2017). To investigate interviewer effects in the present study, the same dependent and independent variables were applied in a multilevel regression. Factor analysis revealed that that the factor *party-political activities* was an appropriate indicator for political participation. An additive index was calculated as a dependent variable measuring political participation. It consisted of the following three items:

- Participation in public discussions at meetings (factor loading: 0.701).
- Participation in a citizens' action group (factor loading 0.697).
- Voluntary work for a political party (factor loading 0.776).

Political anomy was measured with a scale consisting of four items that were summarized into an index that served as a third dependent variable (ZA & ZUMA, 2014). The items were:

- In spite of what some people say, the situation of the average man is getting worse, not better.
- It‘s hardly fair to bring a child into the world with the way things look for the future.
- Most public officials are not really interested in the problems of the average man.
- Most people don‘t really care what happens to the next fellow.

Healthy eating behavior was measured with the question: "On how many days per week do you eat healthy?" to analyze interviewer effects. I have used this variable in the past to explore the impact of falsifications on substantial findings in social science research on the basis of the theory of planned behavior (Landrock & Menold, 2016).

4.2 Independent Variables on the Individual Level

To implement multilevel regression models, statistically significant explanatory variables on the individual level were identified by estimating OLS regressions. These individual-level independent variables were included in the multilevel regression analyses presented in what follows. Given that my research interest here was to estimate interviewer effects, these variables may be considered as control variables.

For *income* as a dependent variable, the statistically significant explanatory variable on the respondent level – besides age – was the living situation, which was measured with the question: "Where are you living during the current semester?" This variable was dichotomized: The option "living with parents or relatives" was coded as 1; other options were coded as 0. The effect of age on income was positive.

Regarding the living situation, the analysis revealed that students who lived with their parents or relatives reported less income than students who did not.

For *political participation,* the statistically significant explanatory variables on the respondent level were internal political efficacy, political dissatisfaction, extremism (captured with the left–right scale), and (female) gender. The means of the individual items were calculated for both internal political efficacy and political dissatisfaction; all items were adapted from the ALLBUS 1998 questionnaire (see Koch et al., 1999).

The items used to measure *internal political efficacy* were:

- I would have the confidence to take on an active role in a group concerned with political issues.
- Politics is so complicated that somebody like me can't understand what's going on at all. (Reverse-scored item)

Political dissatisfaction was measured with the following three items:

- Only when differences in income and social status are large enough is there any incentive for personal achievement.
- Differences in social position between people are acceptable because they basically reflect what one has made of the chances one had.
- I consider the social differences in this country to be just on the whole.

To measure extremism, the left–right scale from the ALLBUS 1998 questionnaire was used:

> Many people use the terms "left" and "right" when they want to describe different political views. Here we have a scale which runs from left to right.
>
> Thinking of your own political views, where would you place these on this scale?

To operationalize extremism (see Lüdemann, 2001), the original 10-point rating scale (with the value 1 on the left end of the scale and the value 10 on the right end of the scale) was recoded in such a way that the original values between 1 and 10 were assigned the new values between 5 and -5. These new values were then squared, thereby yielding a measurement for extremism where the value 1 stands for a very small degree of extremism and the value 25 for a very high degree of extremism (integrating both the left and the right ends of the left–right scale). All of these variables, except extremism, were found to have significant positive effects in the real data. As extremism had a significant positive effect in the falsified data, this independent variable was nonetheless included in the analysis of interviewer effects (Landrock, 2017).

For the dependent variable *political anomy*, two statistically significant explanatory variables, economic dissatisfaction and external political efficacy were identified. *Economic dissatisfaction* was measured with the question: “How would you generally rate the current economic situation in Germany?”

External political efficacy was measured with two items:

- Politicians don’t care much about what people like me think. (Reverse-scored item)
- In general, politicians try to represent the people’s interests.

Here, too, all items were adapted from the ALLBUS 1998 questionnaire. To operationalize external political efficacy, the means of the items were calculated (see Koch et al., 1999). Economic dissatisfaction was found to have a positive influence on political anomy, whereas external political efficacy had a negative effect.

To analyze interviewer effects on reported *healthy eating behavior*, a model based on the theory of planned behavior was adopted, which I applied in previous research on the impact of falsified data on substantive findings (Landrock & Menold, 2016).

The statistically significant independent variables for explaining healthy eating behavior on the individual level are the intention to eat healthily, perceived behavioral control, TV consumption, body mass index, doing sports, and preferring healthy desserts. The intention to eat healthily and perceived behavioral control were measured with two items each. These items were used to calculate an index for intention and for perceived behavioral control:

- In future I will eat healthy at least four days a week. (Intention)
- In the coming weeks I will eat healthy at least four days a week. (Intention)
- It is possible for me to eat healthy at least four days a week. (Perceived behavioral control)
- It is completely in my own hands to eat healthy at least four days a week. (Perceived behavioral control)

The questionnaire included the following question on TV consumption:

> Thinking about the days when you watch TV, how long on average do you watch TV on these days – I mean in hours and minutes?

Body mass index was calculated on the basis of the self-reported height and weight of respondents. The variable *doing sports* was dichotomized; respondents were asked to answer an open-ended question about which sports they took part in at least occasionally. A list of 12 desserts was used to find out whether the respondents preferred healthy desserts. The variable *preference for healthy desserts* was dichotomized. Healthy desserts (fruit curd, fruit salad, or yoghurt) were coded as 1; unhealthy desserts (mousse au chocolat, tiramisu, chocolate pudding, or pancakes) as 0.

As theory-driven explanatory variables, the intention to eat healthily and perceived behavioral control were found to have positive effects on reported healthy eating behavior. TV consumption and body mass index had negative effects, whereas doing sports and preferring healthy desserts showed positive effects, at least in the falsified data.

4.3 Independent Variables on the Contextual Level

One aim of the present study was to identify interviewer characteristics on the contextual level that are linked to interviewer effects. The independent variables on the interviewer level that were tested are variables that are known to generally cause interviewer effects (see hypotheses in section 2 above). These variables are the *payment scheme* (payment per hour vs. payment per completed interview), the *interviewer's gender* and *income*, the *interviewer's response to the same question of the questionnaire*, and the *interviewer's experience*. Interviewers' personality traits were also tested, in particular *extraversion*, *conscientiousness*, and *perceived self-efficacy*, as they were considered relevant for analyzing interviewer effects.

First, the payment scheme was analyzed to determine whether the fact that an interviewer was paid per completed interview or per hour made a difference for the collected data, and therefore for the data quality. Winker et al. (2015) reported such an influence of the payment scheme on formal, non-content-related meta-indicators, for example non-differentiation. The payment scheme was varied in the research design: One half of the interviewers were paid per hour, the other half were paid per completed interview (see also section 3 above).

Many authors have described the core sociodemographic characteristics, namely gender, age, and education, as factors influencing interviewer effects (see Haunberger, 2006; Liu & Stainback, 2013). To my knowledge, researchers usually obtain only this basic information about interviewers from the fieldwork agencies, so that further interviewer characteristics typically cannot be analyzed. In the present study, I included the effects of the interviewers' gender as collected with the questionnaire completed by the interviewers themselves as respondents. Regarding age and education, the data show only small variances because all the interviewers were students and they were therefore very similar with respect to age and education. Instead, I considered the income of the interviewers, assuming that, in the case of the student population of interviewers, income would be an appropriate indicator for the socioeconomic background of an interviewer, which might lead to interviewer effects.

As mentioned above, the interviewers themselves also completed the survey questionnaire as respondents. Thus it was possible to include as an independent variable their responses to the same questions that the respondents were also asked. The interviewers' responses were included as an explanatory variable on the contextual level in order to test whether there were positive correlations between the respondents' answers and the interviewers' answers. Schanz (1981) reported positive correlations between the attitudinal and behavioral characteristics of interviewers and respondents.

A further relevant factor for the occurrence of interviewer effects is interviewer experience (Olson & Bilgen, 2011). The question used to measure this variable was whether the interviewer had ever conducted interviews before participating in the present study. The variable was dichotomized into interviewers with experience and interviewers without experience.

The questionnaire also included scales to measure the personality traits of the interviewers. To analyze the effects of the interviewers' personality traits on the respondents' responses, these traits were included in the multilevel analyses on the contextual level. Perceived self-efficacy was measured as agreement with the following three items (Beierlein, Kovaleva, Kemper, & Rammstedt, 2014) using a seven-point rating scale:

- I can rely on my own abilities in difficult situations.
- I am able to solve most problems on my own.
- I can usually solve even challenging and complex tasks well.

Afterwards, the means of the items were calculated.

To measure extraversion and conscientiousness, the ten-item Big Five Inventory (BFI-10; Rammstedt, Kemper, Klein, Beierlein, & Kovaleva, 2014) with a five-point rating scale was used:

I see myself as someone who...

- ...is reserved (Extraversion, reverse-scored item)
- ...is outgoing, sociable (Extraversion)
- ...tends to be lazy (Conscientiousness, reverse-scored item)
- ...does a thorough job (Conscientiousness)

For these variables, too, the means of each item were calculated.

4.4 Multilevel Regression Model

To test the hypotheses and to investigate whether the interviewers' characteristics influenced the respondents' answers (e.g., reported income), separate identical multilevel regression models were developed for the real and the falsified data. The statistical software Stata 12 was used to conduct the multilevel analyses. First, a null model without an independent variable and without the contextual level was estimated in order to assess the goodness of fit of the baseline model on the basis of log likelihood, or deviance (Hox, 1995). Second, to estimate interviewer-level variance the contextual level was included in the random-intercept-only model (RIOM) in order to be able to answer questions such as whether the income reported by the respondent depended on the interviewer – in other words, whether the incomes of the respondents varied across interviewers. To this end, the intraclass correlation (ICC), which measures interviewer-level variance, was calculated. In the third step, the random-intercept model (RIM) was estimated. This model considers the influence of the individual respondent-level explanatory variables and controls for the contextual level. By including the interviewer-level explanatory variables of the contextual level (intercept-as-outcome model), direct effects of certain interviewer characteristics on respondents' responses were estimated. Thus, it could be determined, for example, whether the income reported by the respondents depended on the interviewers' gender. The results of the intercept-as-outcome model are shown in detail in Tables 4 and 5 (section 5.2).[5]

The likelihood-ratio test and McFadden's R-squared values were used to assess the goodness of fit of the model. With the likelihood-ratio tests, it was assessed, first, whether the multilevel approach was more appropriate than an OLS regression and, second, whether the estimated model extension (i.e., the reduction of deviance) was significant. McFadden's R-squared assesses model fit by comparing the log likelihood of the null model (i.e., the model without dependent variables and contextual level) with the log likelihood of the estimated model. According to Langer (2010, p. 756), values between 0.2 and 0.4 are excellent.

The dependent variables to be analyzed were required to be metric variables. Prior to the analyses, the independent variables were modified: The independent metric variables were grand-mean centered; the independent nominal variables were dichotomized and coded into binary variables.

5 Results

5.1 Interviewer Effects in Real Data

First, interviewer effects in the real data were analyzed. Table 2 shows the random-intercept-only model (RIOM) for all of the dependent variables.[6] The intraclass correlations varied between 0.017 and 0.067, which means that between 1.7 percent and 6.7 percent of the total variance is accounted for by the contextual level (i.e., the interviewer level). These interviewer effects are very small. Only healthy eating behavior, with an ICC of 0.067, showed slightly increased interviewer effects (see Groves & Magilavy, 1986; Mangione et al., 1992). The likelihood-ratio test measures the significance of the models and indicates whether a multilevel model

5 As an extension of the intercept-as-outcome models, the slope-as-outcome models were also estimated; they were not significant.

6 Regarding political anomy, it should be mentioned that there were a large number of missing values, due, in particular, to the item "Most public officials are not really interested in the problems of the average man" (56 missing values).

Table 2 Interviewer effects in the *real data* (random-intercept-only models, RIOMs)

	Dependent Variables			
RIOMs	Income σ^2 (SE)	Political Particip. σ^2 (SE)	Political Anomy σ^2 (SE)	Healthy Eating σ^2 (SE)
Resid. variance (respondents)	143206.6 (8553.957)	0.131 (0.007)	1.183 (0.071)	2.933 (0.165)
Resid. variance (interviewers)	3660.958 (3674.375)	0.002 (0.003)	0.063 (0.034)	0.210 (0.087)
ICC	0.025	0.017	0.050	0.067
LR test (p)	0.1356	0.1834	0.0114	0.0007
N	644	710	623	710

is more suitable than an OLS regression model. Regarding the dependent variables income and political participation, the RIOMs were not significant, which means that multilevel models were not appropriate and OLS regressions should be estimated instead. Regarding political anomy and healthy eating behavior, the RIOMs were significant; multilevel models could thus be preferred over OLS models. In the next step, the individual respondent-level variables were included in the model, and the random-intercept model (RIM) was developed. In the case of political anomy and healthy eating behavior as dependent variables, these models were not significant. Thus it can be assumed that interviewer effects scarcely exist in the real data.

5.2 Interviewer Effects in Falsified data

In the second step, interviewer effects in the falsified data were analyzed accordingly.[7] Table 3 shows the results of the RIOMs. The likelihood-ratio tests indicated that the models for all dependent variables were significant, which implies that the multilevel approach was more appropriate than the OLS regression approach. With values between 0.17 and 0.21, the intraclass correlations were much higher than in the real data, which means that the contextual level explained between 17 and 21% of the total variance. These strong interviewer effects indicate that individual characteristics, attitudes, and behaviors of the interviewers found their way into the

7 In the falsified data, there were a large number of missing cases in the case of income . I assume that the question is difficult to falsify and that the falsifiers therefore preferred to report item nonresponse.

Table 3 Interviewer effects in the *falsified data* (random-intercept-only models, RIOMs)

	Dependent Variables			
RIOMs	Income σ2 (SE)	Political Particip. σ2 (SE)	Political Anomy σ2 (SE)	Healthy Eating σ2 (SE)
Resid. variance (respondents)	30678.33 (1887.241)	0.102 (0.006)	1.125 (0.065)	1.869 (0.105)
Resid. variance (interviewers)	7913.874 (1964.437)	0.020 (0.005)	0.271 (0.065)	0.506 (0.115)
ICC	0.205	0.165	0.194	0.213
LR test (p)	0.0000	0.0000	0.0000	0.0000
N	606	708	681	710

falsified data. Thus, interviewer effects in the falsified data were further analyzed in order to determine which interviewer characteristics, attitudes or behaviors were particularly associated with interviewer effects.

In the third step, the RIOM was extended by including the respondent characteristics on the individual level (RIM, not shown here). Afterwards, the interviewer characteristics on the contextual level were included, thus developing the intercept-as-outcome model (IOM), which estimates the direct effects of the independent variables on the interviewer level. The further extensions of the IOM were not significant for any of the dependent variables. Therefore, the random-intercept, random-slope models with cross-level interactions could not be estimated. Table 4 shows the results of the final IOM for the dependent variables income and political participation.

As can be seen from Table 4, the models fit well: The likelihood-ratio test indicated that both the models themselves and the model extensions to IOMs were significant. The McFadden R-squared values of 0.16 and 0.64 were at least very reasonable.

The results show that all individual variables on the respondent level were significant, at least at the ten percent level, which is not surprising as they already proved to have significant influence in the previously performed OLS regressions. However, for the analysis of interviewer effects, the more relevant results were found on the contextual level. Significant effects on the dependent variables were not found for the payment scheme, the interviewers' personality traits, or the interviewers' experience. The interviewers' income had no significant effect on reported political participation. However, for income and political participation as dependent variables, significant effects of the interviewers' gender and their answers to the same survey questions could be identified.

Table 4 Results of ML regression in the *falsified data* (intercept-as-outcome models, IOMs)

	Dependent Variables			
IOMs	Income		Polit. Particip.	
Fixed Part	Coeff.	SE	Coeff.	SE
Constant	725.907 ***	23.732	0.266 ***	0.036
Respondent level				
Age	10.381 ***	2.345	-	-
Living with parents/ relatives (ref.: no)	-176.879 ***	21.467	-	-
Internal political efficacy	-	-	0.128 ***	0.011
Political dissatisfaction	-	-	0.034 +	0.019
Gender (ref.: m)	-	-	0.035 +	0.019
Extremism	-	-	0.017 ***	0.003
Interviewer level				
Payment per hour (ref.: per int.)	2.435	23.428	-0.025	0.035
Gender (ref.: m)	-51.359 +	26.539	0.086 *	0.039
Income	-	-	0.000	0.000
Interviewer's answer	0.114 *	0.053	0.259 ***	0.052
Experience (ref.: no)	-4.696	29.644	-0.034	0.044
Extraversion	-1.050	14.651	0.017	0.022
Conscientiousness	17.575	15.002	0.022	0.022
Perceived self-efficacy	2.372	12.341	-0.013	0.019
Random Part	σ^2	SE	σ^2	SE
Respondents' residual variance	26933.240	1797.859	0.074	0.005
Interviewers' residual variance	4784.561	1509.125	0.010	0.003
Model fit				
Log likelihood	-3392.254		-92.393	
N	516		579	
LR test (p)	0.0000		0.0000	
LR test model extens. (p)	0.0000		0.0000	
McFadden's R^2	0.1641		0.6433	

Notes: *** p<0.001; ** p<0.01; * p<0.05; + p<0.10

Female falsifying interviewers tended to report lower incomes and higher values for political participation of the respondents than did male falsifying interviewers. Evidence was found that the gender of the interviewer tended to affect reported income and political participation in the case of the falsified data. It was also found that the interviewers' answers to the same questions had a positive effect on the reported respondents' answers. Thus, there were positive correlations between the falsifiers' attitudes and behaviors and the falsified reported attitudes and behaviors of the respondents. Presumably, the interviewers used their own income and political participation as a knowledge base for what a realistic income and political participation level might be for the interviews they were falsifying.

The models estimated for political anomy and healthy eating behavior as dependent variables yielded very similar results (Table 5). In both cases, the interviewers' answers to the same questions had a positive effect on the falsified reported answers of the respondents. In the case of healthy eating behavior as a dependent variable, the interviewers' gender affected the reported falsified response. Male falsifiers reported higher values for healthy eating. Thus, an impact of the attitudes and behaviors of the falsifying interviewers on all four analyzed variables could be identified.

Table 5 Results of ML regression in the *falsified data* (intercept-as-outcome models, IOMs)

	Dependent Variables			
IOMs	Polit. Anomy		Healthy Eating	
Fixed Part	Coeff.	SE	Coeff.	SE
Constant	1.691 ***	0.130	4.580 ***	0.140
Respondent level				
External political efficacy	-0.544 ***	0.045	-	-
Economic dissatisfaction	0.091	0.079	-	-
Intention	-	-	0.353 ***	0.032
Perceived behavioral control	-	-	0.359 ***	0.046
TV consumption	-	-	-0.003 **	0.001
Doing sports (ref.: no)	-	-	0.117 +	0.070
Preference for health desserts (ref.: no)	-	-	0.005	0.010
BMI	-	-	-0.100 ***	0.018
Interviewer level				
Payment per hour (ref.: per interview)	0.041	0.126	-0.089	0.133
Gender (ref.: m)	-0.229	0.148	-0.341 *	0.147
Income	0.000	0.000	0.000	0.000
Interviewer's answer	0.195 ***	0.054	0.160 ***	0.039
Experience (ref.: no)	-0.026	0.156	0.020	0.163
Extraversion	0.085	0.079	0.123	0.083
Conscientiousness	0.032	0.081	0.130	0.083
Perceived self-efficacy	-0.079	0.067	-0.079	0.070
Random Part	σ^2	SE	σ^2	SE
Respondents' resid. variance	0.896	0.057	0.998	0.063
Interviewers' resid. variance	0.133	0.042	0.143	0.047
Model fit				
Log likelihood	-797.383		-827.605	
N	565		565	
LR test (p)	0.0000		0.0000	
LR test model extension (p)	0.0000		0.0000	
McFadden's R^2	0.2613		0.3703	

Notes: *** p<0.001; ** p<0.01; * p<0.05; + p<0.10

5.3 Summary and Review of Hypotheses

First, I will review the two general hypotheses:

H1: Interviewer effects occur both in real and in falsified data.

This hypothesis cannot be confirmed. Interviewer effects were identified in the falsified data but not in the real data.

H2: The interviewer effects in falsified data are larger than in real data.

This hypothesis can be clearly confirmed. Large interviewer effects occurred in the falsified data, whereas interviewer effects could not be identified in the real data.

Next, I will review the more specific hypotheses regarding characteristics of the interviewers that may cause interviewer effects:

H3a: The core sociodemographic characteristics of the interviewers affect the reported responses.

As no effects of the core sociodemographic characteristics of the interviewers were measurable in the real data, this hypothesis must be rejected for the real data. With regard to the falsified data, the analysis of the effect of the interviewers' gender on the dependent variables revealed that female falsifiers reported lower income, higher political participation, and lower values for healthy eating behavior than did their male counterparts. The interviewers' age and education were too homogeneous to be tested. With the exception of income as a dependent variable (see H3c), the interviewers' income does not appear to have affected the falsified responses. Accordingly, for the falsified data, the hypothesis can be confirmed with respect to gender.

H3b: The magnitude of interviewer effects depends on the interviewer's experience.

This hypothesis could not be confirmed for the real or the falsified data: No effect of interviewer experience on any of the dependent variables was found.

H3c: Associations exist between the behaviors and attitudes of interviewers and the reported behaviors and attitudes of the respondents they interview.

This hypothesis cannot be confirmed for the real data, where no interviewer effects were found. However, strong evidence was found in support of the hypothesis in the falsified data: For all four dependent variables, significant positive correlations were found between the interviewers' answers as respondents and the falsified answers to the same survey questions.

H3d: The occurrence and magnitude of interviewer effects depends on the personality traits of the interviewer.

This hypothesis cannot be confirmed for the real data or for the falsified data. No effects of the personality traits on the dependent variables could be identified either in the real data or the falsified data.

H3e: The magnitude of interviewer effects depends on the interviewer payment scheme used (payment per completed interview vs. payment per hour).

This hypothesis cannot be confirmed for the real data or for the falsified data. Although previous research (see Winker et al., 2015) has shown that the payment scheme used (payment per completed interview vs. payment per hour) generally has an impact on the collected data, the present analyses did not detect effects of the payment scheme.

In summary, it can be stated that no interviewer effects of any kind were found in the real data. In the falsified data, the occurrence and magnitude of interviewer effects does not appear to have depended on the interviewers' experience or personality traits, or on the payment scheme used. However, effects of the interviewers' gender were found on the falsified reported income, political participation, and eating behavior of respondents. Furthermore, the interviewers' own attitudes and behaviors were correlated with the falsified reported attitudes and behaviors of the respondents. Thus, the falsifiers' attitudes and behaviors found their way into the falsified data and influenced the data reported as answers of the respondents.

6 Conclusions and Recommendations

The findings of the present study suggest that interviewer effects are clearly stronger in falsified data than in real data: The real data, derived from actual conducted interviews, does not appear to be contaminated by interviewer effects at all. This can be taken as an indication of high data quality, which may be due to the fact that the real interviews were audio-recorded and the fieldwork was intensively monitored. By contrast, very strong interviewer effects were measured in the falsified dataset. This suggests that the process of falsifying leads to a pronounced impact of the falsifiers' sociodemographic characteristics, attitudes, and behaviors on the data reported as answers of the respondents.

However, the interviewer effects (or, more precisely, "falsifier effects") identified in the falsified data were smaller than expected. One reason for this may be that both the respondents and the interviewers were students. Therefore, the falsifiers were familiar with the respondents' social reality and were able to give realistic answers – which reduced the magnitude of the interviewer effects. (This may also be a reason for the absence of interviewer effects in the real data.) A second reason why interviewer effects in the falsified data were smaller than expected may be that, despite the fact that the dependent variables used were empirically shown to be susceptible to interviewer effects, more appropriate dependent variables could possibly have been found to analyze interviewer effects.

The fact that neither the payment scheme nor the interviewers' experience caused interviewer effects is surprising because current findings in the literature suggest that they should have. Winker et al. (2015) found that the payment scheme had an impact on formal, non-content-related meta-indicators such as non-differentiation. However, the present study analyzed content-related dependent variables. A further reason why the payment scheme did not have the hypothesized influence could be that the instructed falsifiers in the present experimental study had an intrinsic motivation to participate in the study and were therefore less frustrated by payment per completed interview than an interviewer in a real fieldwork setting might have been. Moreover, the interviewers in the present study selected the respondents on the university campus and interviewed them themselves. In a real fieldwork setting, the interviewers must contact certain predefined target persons, which may be time-consuming. In such a case it would appear plausible that the payment scheme would make a difference and that payment per hour might enhance motivation to contact the predefined target person. The lack of support for the hypothesized influence of interviewer experience might be due to the fact that the students who stated that they had conducted interviews before were still less experienced than the experienced interviewers in the studies in which interviewer effects have been found.

One limitation of the present study is the fact that the respondents and interviewers were students and that core sociodemographic characteristics, such as age and education, displayed only small variance. Moreover, in a real fieldwork setting, it would hardly be possible to implement an experimental approach such as that employed here. Nonetheless, I assume that the present results are generalizable, not least because interviewers in social science research and market research are often students. However, further research will be needed to confirm the generalizability of my results to real survey settings.

A number of recommendations can be derived from the present findings. First, researchers conducting interviewer-based surveys should collect as much information about the interviewers as possible and feasible (see Bogner & Landrock, 2016; Winker et al., 2015). In particular, as the present study shows, interviewer responses to the same questions that the respondents are asked are highly suitable for detecting interviewer effects in the case of falsified interviews. The interviewers could be requested to complete the survey questionnaire as part of interviewer training, for example. This would have at least two positive effects: First, the interviewers would familiarize themselves with the questionnaire, as a preparation for conducting the interviews; second, the researchers could get to know the interviewers.

A further recommendation that can be derived from the findings of the present study is that researchers using interviewer-based data should check the data for interviewer effects, especially if they suspect that falsifications may have occurred. Falsification checking should be implemented at least by calculating intraclass correlations or conducting multilevel analyses as presented in this paper. This can be done for the entire dataset or only for suspicious cases – provided, of course, more than one interviewer is involved. If a large share of the variance is explained by interviewer-level variables, this may be an indication of contamination of the dataset by interviewer falsifications. In light of the fact that neither bivariate nor multivariate correlational analyses have proved effective in unambiguously establishing the existence of falsifications, the assessment strategies presented here may be very valuable for improving the quality and accuracy of survey data.

References

Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2014). Allgemeine Selbstwirksamkeit Kurzskala (ASKU). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis35. Retrieved from http://zis.gesis.org/skala/Beierlein-Kovaleva-Kemper-Rammstedt-Allgemeine-Selbstwirksamkeit-Kurzskala-(ASKU)

Beullens, K., & Loosveldt, G. (2016). Interviewer effects in the European Social Survey. *Survey Research Methods*, *10*(2), 103-118. doi:10.18148/srm/2016.v10i2.6261.

Bogner, K., & Landrock, U. (2016). Response biases in standardised surveys. *GESIS Survey Guidelines*, Mannheim, Germany: GESIS – Leibniz Institute for the Social Sciences. doi: 10.15465/gesis-sg_en_016.

Groves, R. M., & Magilavy, L. J. (1986). Measuring and explaining interviewer effects in centralized telephone surveys. *Public Opinion Quarterly*, *50*(2), 251-266.

Haunberger, S. (2006). Das standardisierte Interview als soziale Interaktion: Interviewereffekte in der Umfrageforschung. *ZA-Information*, 58, 23-46.

Hox, J. J. (1995). *Applied multilevel analysis*. Amsterdam: TT-Publikaties.

Koch, A., Kurz, K., Mahr-George, H., & Wasmer, M. (1999). Konzeption und Durchführung der “Allgemeinen Bevölkerungsumfrage der Sozialwissenschaften” (ALLBUS) 1998. Mannheim: *ZUMA-Arbeitsbericht* 1999/02. Retrieved from http://nbn-resolving.de/urn:nbn:de:0168-ssoar-200413

Landrock, U. (2017). Explaining political participation: A comparison of real and falsified survey data. *Statistical Journal of the IAOS*, *33*(2), 447-458. doi: 10.3233/SJI-160270.

Landrock, U., & Menold, N. (2016). Validation of theoretical assumptions with real and falsified survey data. *Statistical Journal of the IAOS*, *32*(3), 305-312. doi: 10.3233/SJI-161020.

Langer, W. (2010). Mehrebenenanalyse mit Querschnittsdaten. In C. Wolf, & H. Best (Eds.), *Handbuch der sozialwissenschaftlichen Datenanalyse* (pp. 741-774), Wiesbaden: VS Verlag für Sozialwissenschaften.

Liu, M., & Stainback, K. (2013). Interviewer gender effects on survey responses to marriage related questions. *Public Opinion Quarterly*, *77*(2), 606-618.

Lüdemann, C. (2001). Politische Partizipation, Anreize und Ressourcen. Ein Test verschiedener Handlungsmodelle und Anschlusstheorien am ALLBUS 1998. In A. Koch, P. Schmidt, & M. Wasmer (Eds.), *Politische Partizipation in der Bundesrepublik*

Deutschland. Empirische Befunde und theoretische Erklärungen (pp. 43–71). Opladen: Leske and Budrich.

Mangione, T. W., Fowler, F. J., & Louis, T. A. (1992). Question characteristics and interviewer effects. *Journal of Official Statistics*, *8*(3), 293-307.

Menold N., & Kemper C. J. (2014). How do real and falsified data differ? Psychology of survey response as a source of falsification indicators in face-to-face surveys. *International Journal of Public Opinion Research*, *26*(1), 41-65. doi: 10.1093/ijpor/edt017.

Olson, K., & Bilgen, I. (2011). The role of interviewer experience on acquiescence. *Public Opinion Quarterly*, *75*(1), 99-114.

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2014). Big Five Inventory (BFI-10). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*, doi:10.6102/zis76. Retrieved from http://zis.gesis.org/skala/Rammstedt-Kemper-Klein-Beierlein-Kovaleva-Big-Five-Inventory-(BFI-10)

Reuband, K.-H. (1990). Interviews, die keine sind. „Erfolge“ und „Mißerfolge“ beim Fälschen von Interviews. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, *42*(4), 706-733.

Schanz, V. (1981). Interviewereffekte: zusammenfassende Darstellung von Arbeiten, die im Rahmen zweier von ZUMA betreuter Projekte entstanden sind. *ZUMA Nachrichten 9*, 36-46.

Schnell, R. (1991). Der Einfluss gefälschter Interviews auf Survey-Ergebnisse. *Zeitschrift für Soziologie*, *20*(1), 25–35.

Schraepler, J.-P., & Wagner, G. G. (2003). Identification, characteristics and impact of faked interviews in surveys. An analysis by means of genuine fakes in the raw data of SOEP. *IZA Discussion Paper No. 969*, Forschungsinstitut zur Zukunft der Arbeit.

West, B. T., & Blom, A. G. (2016). Explaining interviewer effects: A research synthesis. *Journal of Survey Statistics and Methodology*, 0, 1-37. doi: 10.1093/jssam/smw024.

Winker P., Kruse, K.-W., Menold, N., & Landrock U. (2015). Interviewer effects in real and falsified interviews: Results from a large scale experiment. *Statistical Journal of the IAOS*, *31*(3), 423–434. doi: 10.3233/SJI-150908.

Zentralarchiv für empirische Sozialforschung (ZA) & Zentrum für Umfragen, Methoden und Analysen (ZUMA) e.V. (2014). Anomie (ALLBUS). *Zusammenstellung sozialwissenschaftlicher Items und Skalen*. doi:10.6102/zis58.Retrieved from http://zis.gesis.org/skala/ZA-ZUMA-Anomie-(ALLBUS)