*Journal of Survey Statistics and Methodology (2014) **2**, 433–458*

INTERVIEWER EFFECTS ON LATENT CONSTRUCTS IN SURVEY RESEARCH

KOEN BEULLENS*
GEERT LOOSVELDT

Interviewers may have an influence on the answers given by their respondents. Survey researchers usually estimate these interviewer effects on the univariate distributions of survey variables. However, it is also possible that associations between survey items are influenced by interviewer effects. In this paper, we analyze the covariance structure for a set of survey items that can be used for factor analysis. It is hypothesized that interviewers affect the covariances between items belonging to the same, as well as to different, latent constructs. As a consequence, factor loadings and correlations between latent constructs may be biased when interviewer effects are ignored. In order to assess the effects of interviewers on latent construct analysis, multilevel covariance analysis is performed on nine items belonging to three latent constructs derived from data of eight different countries in the European Social Survey, round 5. Results indicate that interviewer effects on these associations occur, but that their impact on measurement models is rather modest.

KEY WORDS: Interviewer effects on covariances; Interviewer variance; Multilevel covariance analysis.

1. INTRODUCTION

Interviewer variance can be seen as the correlated responses of respondents interviewed by the same interviewer. A usually relatively small but substantive part of the variance of respondents' answers can be explained by interviewer clustering. Intraclass correlations (ICCs) that are found throughout the relevant literature range roughly between 0.00 and 0.05, sometimes increasing to 0.10, with some outliers exceeding 0.10 (see, among others, Kish 1962; Freeman

KOEN BEULLENS is with the Centre for Sociological Research, KU Leuven, Parkstraat 45, Box 3601, Leuven, Belgium. GEERT LOOSVELDT is with the Centre for Sociological Research, KU Leuven, Parkstraat 45, Box 3601, Leuven, Belgium.
*Address correspondence to Koen Beullens; e-mail: koen.beullens@soc.kuleuven.be.

doi: 10.1093/jssam/smu019 Advance access publication 4 October 2014

and Butler 1976; Mangione, Fowler, and Louis 1992; Groves and Magilavy 1986; O'Muircheartaigh and Campanelli 1998).

Interviewer variance is not necessarily the same as interviewer bias (Biemer and Lyberg 2003; Loosveldt 2008). Interviewer bias results from dominant and systematic effects of all interviewers on the obtained answers. In this respect, social desirability is a good example. The presence of an interviewer in itself systematically directs the respondents' answers, as the presence of an interviewer activates social norms in the answering process. Some characteristics of interviewers, such as gender or race, can also give rise to different tendencies in respondents' answers.

Usually, interviewer variance is interpreted as a source of survey error related to measurement error, as interviewers are prone to affecting the responses of their respondents. As interviewers behave differently during an interview (e.g., probing), they create a particular atmosphere, which may affect the expectations of respondents or the role they need to play (Mangione et al. 1992). This will probably affect attitudinal questions more than factual questions (Schnell and Kreuter 2005).

In general, it is believed that when interviewers do their job in a standardized way and adhere to the interview rules, they will obtain comparable answers provided that the groups of respondents they are assigned are also comparable (Loosveldt 2008). However, since interviewers in face-to-face surveys are usually assigned to respondents who live relatively close to the interviewer, there is always a risk that interviewer and area effects cannot be disentangled.

Differences between interviewers may indeed be explained simply by the fact that they are assigned to different sets of addresses or sample units. Interviewers working in an urban environment may be dealing with different types of respondents, compared with more rural-oriented interviewers. Disentangling these area effects from measurement error can be achieved only by restricting the fieldwork to an interpenetrated design (O'Muircheartaigh and Campanelli 1998; Schnell and Kreuter 2005). Interpenetrated designs (Mahalanobis 1946) can remove such a confounding by assigning respondents at random to interviewers. Recently, West, Kreuter, and Jaenichen (2013) and West and Olson (2010) have also argued that interviewer effects may originate from differences in interviewers' (non)response profiles. As some interviewers are better at obtaining responses from, for example, women, foreigners, or more reluctant participants, they produce interviews that come from a particular type of respondent. This may in turn lead to correlated answers from the respondents recruited by the same interviewer.

Whatever the source of interviewer effects may be, the resulting clustering has an unfavorable (design) effect on the survey estimates (Kish 1962), depending on the size of the intra-interviewer correlation and the average interviewer workload. In face-to-face surveys, such effects may even double the standard errors of the estimates (Schnell and Kreuter 2005; Loosveldt and Beullens

2010). These interviewer effects influence not only location estimates, such as averages or proportions, but also associations between variables, though to a lesser extent than location parameters (Beullens and Loosveldt 2013).

In this regard, Davis and Scott (1995) (as cited by Rao [2005]) discuss effects of interviewers on domain comparisons. They find that the effect of interviewer variability on the response variance is smaller when interviewers recruit respondents from two domains. Such domain comparisons can be seen as a form of bivariate association.

The fact that the correlations between two survey variables can be altered by interviewer effects may particularly apply to survey questions that relate to the same underlying latent construct. For example, suppose that Interviewer 1 has a slight tendency to push the answers to the survey questions of his or her respondents toward the negative end of the scale, whereas Interviewer 2 instead pushes the answers toward the positive end. As a result of these interviewer effects, the overall correlations between the items of the same latent construct might be somewhat exaggerated.

Figure 1 illustrates how such a problem can be observed, given a three-level hierarchical data structure. Three items of the same latent construct are used here. The answers of the respondents to each of these items are located at the lowest level, nested within the second level of respondents (ID1 to ID10), who are in turn nested within individual interviewers (Int1 and Int2). The entire variability of the answers to the items y_i of the same latent construct can be attributed to any of the three levels:

$$y_{ijk} = \alpha_i + \mu_j + \nu_k + \varepsilon_{ijk}.$$

Each of the items $i = 1, 2, 3$ is allowed to have its own mean or intercept α_i, and individual respondents $j = 1, 2, \ldots, J$ can express their particular deviation μ_j from the overall mean. It is preferable for the interviewer $k = 1, 2, \ldots, K$ not to show substantial deviations ν_k on the different items. Therefore, relative to the variance of the respondents σ_J^2, the interviewer variance σ_k^2 should be small. Unfortunately, as interviewers may have a tendency to push the

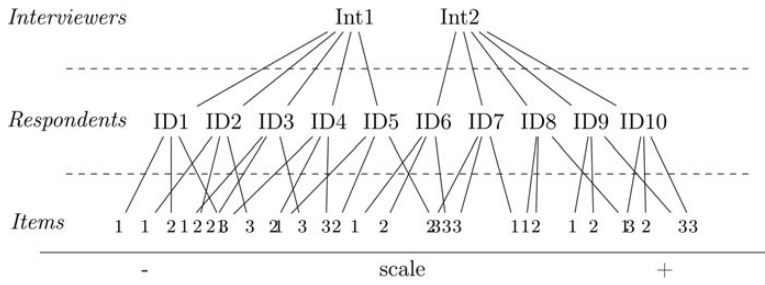

Figure 1. Three-Level Structure of the Items of a Latent Construct.

answers of the respondents toward the positive or negative end of the scale, some proportion of the total variance of the items in the construct is interviewer variance, even after taking the respondent variance σ_j^2 into account.

In the fictitious illustration shown in figure 1, the level 2 intra-respondent correlation is very high, as the different measurements of the items are located very close to one another within each respondent. At level 3, the intra-interviewer correlation also seems to be substantial, as all measurements for Interviewer 1 are located on the negative side of the scale, while those for Interviewer 2 are situated on the positive side of the scale. Because of the intra-interviewer correlation and the intra-respondent correlations, the three survey items may be considerably correlated, although the correlations may be overestimated because of interviewer effects.

The same problem can also be interpreted in terms of a two-level factor analysis. Usually, (confirmatory) factor analysis is informed by the covariance matrix of a set of survey items, which are assumed to be explained by one or more underlying latent concepts. However, if there is reason to believe that some part of the elements of the covariance matrix are overestimated because of interviewer variance, simply using the total covariance matrix may introduce bias in the factor loadings, factor variances, and even the covariances between the assumed latent variables. Therefore, it may be worthwhile decomposing the total covariance matrix $\mathbf{\Sigma_T}$ into a within-interviewer covariance matrix $\mathbf{\Sigma_W}$ and a between-interviewer covariance matrix $\mathbf{\Sigma_B}$. Ideally, there are no interviewer effects, in which case $\mathbf{\Sigma_B} = 0$.

It is hypothesized in this paper that between-interviewer effects do not apply only to variances (diagonal elements on the covariance matrix $\mathbf{\Sigma_B}$), but that they also affect its off-diagonal elements, possibly biasing the results of the measurement of the latent constructs.

Interviewer effects on the means or proportions of the target variables of a survey have been monitored many times. Now, this paper seeks to extend research on interviewer effects to multivariate statistics, particularly elaborating on the case of a measurement model in which three latent sociological concepts are presented. This way, interviewer effects are more integrated into analytical statistics or statistics of substantive interest. Therefore, this paper supports the call of the *Journal of Survey Statistics and Methodology* that "more research is needed to improve inferences for 'analytical' purposes—that is, where the objective is to explore relationships among variables rather than simple description" (Sedransk and Tourangeau 2013, pp. 4–5).

2. DATA

Nine survey items from the fifth round of the European Social Survey (ESS-5) are used in the analyses, representing three latent constructs. The three latent constructs are "Social trust," "Political trust," and "Perceived threat from

immigrants," each derived from three questions with all items measured on a scale from 0 to 10 points. The three sets of survey questions are all located early in the questionnaire, and all belong to the core module of the ESS-5 questionnaires.

The three "Social trust" items are the eighth, ninth, and tenth questions in the questionnaire of ESS-5. These items have been presented to the respondents as follows:

You can't be too careful — **Most people can be trusted**

0 1 2 3 4 5 6 7 8 9 10

Most people would try to take advantage of me — **Most people would try to be fair**

0 1 2 3 4 5 6 7 8 9 10

People mostly look out for themselves — **People mostly try to be helpful**

0 1 2 3 4 5 6 7 8 9 10

These three items will be further referred to as PPLTRST, PPLFAIR, and PPLHLP.

The "Political trust" items are at positions 14, 17, and 18 of the questionnaire. These items refer to the degree of trust people feel toward their national parliament (TRSTPRL), the politicians (TRSTPLT), and political parties (TRSTPRT) in their country. Respondents could answer these three items by using the following response scale:

No trust at all — **Complete trust**

0 1 2 3 4 5 6 7 8 9 10

The items about immigrants are situated slightly further on than the items about political trust, in question positions 48, 49, and 50. The questions refer to the effects the immigrants have on three different dimensions of the respondents' country, as indicated by the three following scales:

Bad for the economy — **Good for the economy**

0 1 2 3 4 5 6 7 8 9 10

Cultural life undermined — **Cultural life enriched**

0 1 2 3 4 5 6 7 8 9 10

Worse place to live — **Better place to live**

0 1 2 3 4 5 6 7 8 9 10

These items will be further referred to as IMBGECO, IMUECLT, and IMWBCNT. These three sets of items are chosen because they each bring together three items per latent construct, use the same 0-to-10-point scale, and are all located early in the questionnaire. Furthermore, as they constitute three different concepts, the covariances between them can also be assessed with regard to interviewer effects.

Eight countries are chosen, of which a first set of four countries used individual-based samples (Finland, Belgium, Spain, and Slovenia). The second set of countries comprises household- or address-based sample frames (the Netherlands, Israel, Lithuania, and Greece). ESS countries using an individual-based sample frame seem to have lower levels of intra-interviewer correlations (Loosveldt and Beullens 2010; Beullens and Loosveldt 2013). A speculative explanation for these differences between individual sample frames, on the one hand, and household- or address-based samples, on the other, is that interviewers may be inclined to select just any person from a household that is most convenient to the interviewer. So, a selection bias may be the cause of these increased interviewer variances in household- or address-based samples as compared to individual-based samples. In both forms of sample frames, interviewers can of course still elicit selection effects through nonresponse.

Based on the first round of the European Social Survey, Philippens and Loosveldt (2004) studied the presence of interviewer-induced variance in more than 20 participating countries. The inter-interviewer variance was determined over 105 survey items. They found that the size of intra-interviewer correlation strongly varies across countries, ranging from 0.05 for the Scandinavian countries to 0.15–0.20 for the Southern European countries. Comparable levels of inter-interviewer variance were also found for rounds 4 and 5 of the ESS (Loosveldt and Beullens 2010; Beullens and Loosveldt 2013).

It may be interesting to use countries showing different levels of intra-interviewer correlations, as this may also predict the susceptibility of a country to produce interviewer-affected covariances between the items, in addition to the univariate interviewer variance on a single survey item. Table 1 shows the medians per ESS country over a range of all (101) attitudinal questions in the ESS-5 questionnaire. The selected countries for further analysis are indicated (bold) in the table. It can be seen that the countries are selected in such a way that low ICC countries as well as high ICC countries are chosen to be analyzed. In table 1, the values between brackets indicate the remaining ICC after respondent-level information has been removed (controlled for). These variables include age, gender, job status, citizenship, degree of urbanization, marital status, and level of education.

Table 2 provides an overview of the intra-interviewer correlations across the eight countries for the items of the three latent constructs. Regarding the individual-based samples, the interviewer effects are clearly relatively small in Finland, followed by respectively Belgium, Spain, and Slovenia—where the ICCs are considerable, even reaching a level of 0.15, For household- or

Table 1. Median Intra-Interviewer Correlations for 101 Attitudinal Survey Questions per Country, ESS-5

(a) Individual-based samples					
Sweden	0.01 (0.01)	Germany	0.06 (0.05)		
Finland	0.02 (0.01)	**Spain**	0.07 (0.07)		
Norway	0.02 (0.02)	Estonia	0.08 (0.07)		
Denmark	0.02 (0.02)	**Slovenia**	0.08 (0.08)		
Belgium	0.04 (0.04)	Poland	0.11 (0.11)		
Switzerl.	0.05 (0.05)	Hungary	0.12 (0.12)		
(b) Household- or address-based samples					
The Netherlands	0.03 (0.02)	Ireland	0.14 (0.14)	Slovakia	0.18 (0.18)
France	0.03 (0.03)	Hungary	0.12 (0.12)	Ukraine	0.25 (0.25)
Czech R.	0.04 (0.04)	Cyprus	0.18 (0.18)	Russia	0.22 (0.22)
Great-Br.	0.05 (0.04)	**Lithuania**	0.16 (0.16)	**Greece**	0.22 (0.23)
Israel	0.13 (0.13)	Portugal	0.19 (0.20)	Bulgaria	0.24 (0.24)

address-based samples, the intra-interviewer correlations seem to be systematically higher as compared to the individual-based samples.

It would also be possible to construct a one-factor model for each latent concept in each country and then determine the ICC based on the factor scores. Per assumed construct and per country, a factor analysis (principal components) was used to construct the new variables containing the factor scores. Table 2 again reflects that Finland would have the lowest degree of ICC, whereas Slovenia would have the highest levels among the individual-based samples, and that the Netherlands has the lowest degree of ICC, whereas Lithuania and Greece have the highest degree of ICC among the household- or address-based samples. The construct reflecting political trust seems to be the least affected by the interviewer effects.

A complicating factor when examining interviewer effects is the possible confounding of interviewer and area effects. This means that part of the interviewer variance σ_k^2 should possibly be attributed to the respondent variance σ_j^2. Therefore, the interviewer effects that are observed act as a "worst-case scenario" benchmark. Nonetheless, as interviewer effects are widely recognized, simply ignoring the problem and solely relying on the total variance may be too naive. As Schnell and Kreuter (2005) found that interviewer effects are larger than area effects, some part of the total variance is due to interviewer variance and consequently should be removed.

In the ESS, interviewers are usually assigned to cases they live relatively close to, in order to save on transportation costs. This makes it hard to disentangle area and interviewer effects. However, table 1 also shows the intra-interviewer correlations after removing or controlling for respondent-level information (ICC's between brackets). These control variables include age, gender, job status,

Table 2. Intra-Interviewer Correlation on Separate Items and Factor Scores, ESS-5

(a) Individual-based sample				
	Finland	Belgium	Spain	Slovenia
Respondents	1878	1703	1885	1403
Interviewers	128	128	67	65
Social trust				
PPLTRST	0.0283	0.0203	0.0654	0.1414
PPLFAIR	0.0154	0.0142	0.0453	0.1448
PPLHLP	0.0169	0.0625	0.1333	0.1416
Factor	0.0177	0.0401	0.0657	0.1341
Political trust				
TRSTPRL	0.0269	0.0570	0.0272	0.0733
TRSTPLT	0.0195	0.0306	0.0302	0.0811
TRSTPRT	0.0215	0.0185	0.0385	0.0756
Factor	0.0120	0.0220	0.0256	0.0420
Perceived threat from immigrants				
IMBGECO	0.0399	0.0319	0.0994	0.1340
IMUECLT	0.0289	0.0437	0.0813	0.1526
IMWBCNT	0.0369	0.0710	0.0940	0.1425
Factor	0.0444	0.0580	0.0892	0.1520
(b) Household- or address-based sample				
	The Netherlands	Israel	Lithuania	Greece
Respondents	1833	2295	1684	2715
Interviewers	158	93	103	139
Social trust				
PPLTRST	0.0440	0.0775	0.1908	0.2176
PPLFAIR	0.0205	0.1331	0.1385	0.1900
PPLHLP	0.0107	0.1277	0.1872	0.2453
Factor	0.0277	0.1356	0.2100	0.2526
Political trust				
TRSTPRL	0.0286	0.0414	0.1716	0.1607
TRSTPLT	0.0115	0.0542	0.1571	0.1310
TRSTPRT	0.0220	0.0688	0.1743	0.1346
Factor	0.0272	0.0561	0.1841	0.1526
Perceived threat from immigrants				
IMBGECO	0.0396	0.1732	0.1681	0.2587
IMUECLT	0.0397	0.2371	0.1932	0.2189
IMWBCNT	0.0448	0.2794	0.2479	0.2139
Factor	0.0505	0.2883	0.2551	0.2567

citizenship, degree of urbanization, marital status, and level of education. Such an intervention makes the groups assigned to different interviewers more comparable, as it is assumed they partially filter out the area effects. The results in table 1 indicate that there are hardly any differences between the raw intra-

interviewer correlations and the conditional ones, suggesting that the area effects are not the dominant sources of interviewer differences.

It is also possible to approach the same data using a three-level random intercepts model. As indicated by figure 1, respondents at the second level are nested in the interviewers at the top level and the individual measurements of the items are at the lowest level. In this three-level perspective, the answers to the three items of the same construct can be considered as repeated measurements. Table 3 shows how the variance of the survey items is proportionally distributed over the different levels of observation. With regard to the "Social trust" concept, between 28 and 66 percent of the variance is residual variance, leaving less than half of the variance at the respondent level. The "Political trust" concept has the lowest residual variance and the highest inter-respondent variance. "Perceived threat from immigrants" takes the middle position. A small but substantive proportion is attributable each time to the interviewer

Table 3. Proportions of Variance for Three Survey Items Decomposed into Interviewer Level σ_k^2, Respondent Level σ_j^2, and Error Variance σ_ε^2, for Three Constructs and Eight ESS-5 Countries

Construct	Variance component				
(a) Individual-based sample					
		Finland	Belgium	Spain	Slovenia
Social trust	Interviewer	0.0179	0.0188	0.0490	0.1120
	Respondent	0.4577	0.3629	0.2915	0.4267
	Residual	0.5251	0.6182	0.6594	0.4612
Political trust	Interviewer	0.0174	0.0278	0.0218	0.0683
	Respondent	0.7630	0.6837	0.6721	0.6787
	Residual	0.2206	0.2885	0.3061	0.2530
Perceived threat from immigrants	Interviewer	0.0326	0.0311	0.0702	0.1151
	Respondent	0.5995	0.5364	0.5553	0.5424
	Residual	0.3696	0.4325	0.3745	0.3425
(b) Household- or address-based sample					
		The Netherlands	Israel	Lithuania	Greece
Social trust	Interviewer	0.0174	0.0360	0.0865	0.1844
	Respondent	0.3712	0.4164	0.4121	0.4117
	Residual	0.6114	0.5476	0.5014	0.4039
Political trust	Interviewer	0.0210	0.0397	0.0459	0.1234
	Respondent	0.7064	0.7350	0.6499	0.5759
	Residual	0.2725	0.2252	0.3042	0.3007
Perceived threat from immigrants	Interviewer	0.0317	0.0707	0.2017	0.2050
	Respondent	0.4380	0.6443	0.4360	0.5068
	Residual	0.5303	0.2850	0.3622	0.2882

level, where Finland shows the lowest proportion of interviewer variance and Greece the highest.

As this three-level model decomposes only the variance of the items, a more extensive model is needed that also allows the decomposition of the respective covariances. These operations are explained further in the next section.

3. MULTILEVEL CONFIRMATORY FACTOR ANALYSIS

Given the nine items and the expectation that they can be explained by three underlying (and possibly mutually correlated) constructs, researchers would usually apply structural equation models. Such an analysis uses the covariance matrix of the nine items and arranges this information into a more interpretable solution, providing factor loadings, residual terms, and covariances between the factors. However, as the diagonal elements of the covariance matrix (the variances of the items)—and probably also the off-diagonal elements reflecting the relationships between the items—may be biased because of interviewer effects, the solution offered by the structural equation model may also have to deal with this bias.

A possible solution to this problem consists of decomposing the total or raw covariance matrix $\mathbf{\Sigma_T}$ into a within-interviewer matrix $\mathbf{\Sigma_W}$ and a between-interviewer matrix $\mathbf{\Sigma_B}$, using multilevel covariance structure analysis (Muthén 1989, 1994). According to Muthén, the total covariance matrix can be decomposed into the between-level and within-level counterparts with respect to a multivariate vector $\mathbf{y}$:

$$\mathbf{S_T} = (N-1)^{-1} \sum_{k=1}^{K} \sum_{j=1}^{N_k} (\mathbf{y}_{kj} - \bar{\mathbf{y}})(\mathbf{y}_{kj} - \bar{\mathbf{y}})',$$

$$\mathbf{S_{PW}} = (N-K)^{-1} \sum_{k=1}^{K} \sum_{j=1}^{N_k} (\mathbf{y}_{kj} - \bar{\mathbf{y}}_k)(y_{kj} - \bar{\mathbf{y}}_k)',$$

$$\mathbf{S_B} = (K-1)^{-1} \sum_{k=1}^{K} N_K (\bar{\mathbf{y}}_k - \bar{\mathbf{y}})(\bar{\mathbf{y}}_k - \bar{\mathbf{y}})',$$

where $\mathbf{S_T}$ is a consistent estimator of the total covariance matrix $\mathbf{\Sigma_B} + \mathbf{\Sigma_W}$ and the pooled within-matrix $\mathbf{S_{PW}}$ is a consistent and unbiased estimator of $\mathbf{\Sigma_W}$. $\mathbf{S_B}$ is a consistent and unbiased estimator of $\mathbf{\Sigma_W} + c\mathbf{\Sigma_B}$, provided that

$$c = \frac{N^2 - \sum_{k=1}^{K} N_k^2}{N(K-1)}.$$

The index $k = 1, 2, \ldots, K$ identifies the interviewers, and c is a constant that approximates the average number of respondents per interviewer. The index $j = 1, 2, \ldots, J$ identifies the respondents.

If it is then assumed that all differences between the interviewers are real interviewer effects (and not spurious area effects), the within-interviewer covariance matrix can be used instead of the total covariance matrix to inform the structural equation model that is of substantive interest.

Table 4 shows the result of the decomposition of the total covariance matrix for the three items attributable to the concept "Social trust" into the within and between counterparts. The lower left triangle shows the covariances, and the upper right triangle shows the respective correlations. The diagonal elements are the variances.

The between-interviewer covariances at the lower left of the $\mathbf{\Sigma_B}$-matrices are particularly interesting. They seem to be relatively small in Finland, Belgium, and the Netherlands, but are considerable in Slovenia, Israel, Lithuania, and Greece. When the between-interviewer (co)variances are substantial, the total covariance structure may also include interviewer effects, which should not be reflected in the factor analysis that researchers are usually interested in. Not only is there evidence of substantial variances on the diagonal of the between-interviewer covariance matrix, but the off-diagonal covariances also seem to be substantial. For example, in Greece (see table 4), the total covariance for PPLTRST is 5.479, whereas the between-variance for that variable is 1.161, indicative for all interviewer-specific deviations from the mean of PPLTRST. This interviewer-specific variance is substantial, as was already found in table 2. The between variance for PPLFAIR is 0.872, which is also considerable relative to the total variance of 4.7. Now, both vectors of interviewer-specific deviation for PPLTRST and PPLFAIR seem to covary (0.917). In other words, as an interviewer systematically finds his respondents to give more positive answers to item PPLTRST, it is likely that these respondents will also give more positive answers to item PPLFAIR.

The more substantial the between-interviewer covariance elements, the more the correlation of the within-interviewer covariance matrix deviates from the total covariance matrix. For Finland, Belgium, and the Netherlands, where the elements of the between-interviewer covariance matrix are relatively small, the correlations shown by the total and within matrices hardly differ. For Slovenia, Lithuania, and Greece, the differences between the total and within correlations are larger, in particular because the elements on the between-interviewer covariance matrix are substantial. Here again, the countries with household- or address-based samples seem to have larger differences between total and within correlations, as could be expected based on tables 1–3.

Given the assumption that the differences between the interviewers are real interviewer effects, the measurement model should not be based on the total covariance structure, but instead on the within-interviewer covariance matrix. In this way, interviewer effects are eliminated from the analysis. In fact, multilevel factor analysis provides a factor analysis at both the within-interviewer and the between-interviewer levels. The essential question is whether there are

Table 4. Covariances (Lower Left) and Correlations (Upper Right) between Items of the Construct "Social Trust," for the Total Covariance Matrix Σ_T, the Within-Interviewer Covariance Matrix Σ_W, and the between-Interviewer Covariance Matrix Σ_B, ESS-5

(a) Individual-based sample

	Finland				**Belgium**		
Σ_T	PPLTRST	PPLFAIR	PPLHLP	Σ_T	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	3.619	*0.534*	*0.457*	PPLTRST	4.383	*0.483*	*0.351*
PPLFAIR	1.809	3.174	*0.435*	PPLFAIR	1.922	3.619	*0.310*
PPLHLP	1.673	1.490	3.698	PPLHLP	1.473	1.182	4.010
Σ_W	PPLTRST	PPLFAIR	PPLHLP	Σ_W	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	3.508	*0.527*	*0.451*	PPLTRST	4.300	*0.486*	*0.335*
PPLFAIR	1.746	3.125	*0.428*	PPLFAIR	1.905	3.570	*0.314*
PPLHLP	1.613	1.445	3.643	PPLHLP	1.354	1.154	3.791
Σ_B	PPLTRST	PPLFAIR	PPLHLP	Σ_B	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	0.109	*0.851*	*0.756*	PPLTRST	0.078	*0.285*	*0.875*
PPLFAIR	0.062	0.048	*0.852*	PPLFAIR	0.018	0.051	*0.310*
PPLHLP	0.058	0.043	0.054	PPLHLP	0.112	0.032	0.211
	Spain				**Slovenia**		
Σ_T	PPLTRST	PPLFAIR	PPLHLP	Σ_T	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	3.781	*0.472*	*0.287*	PPLTRST	5.928	*0.622*	*0.487*
PPLFAIR	1.689	3.386	*0.261*	PPLFAIR	3.716	6.020	*0.533*
PPLHLP	1.151	0.991	4.251	PPLHLP	2.809	3.099	5.607
Σ_W	PPLTRST	PPLFAIR	PPLHLP	Σ_W	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	3.538	*0.456*	*0.278*	PPLTRST	5.134	*0.580*	*0.422*
PPLFAIR	1.544	3.239	*0.263*	PPLFAIR	2.989	5.163	*0.480*
PPLHLP	1.006	0.911	3.695	PPLHLP	2.099	2.391	4.813
Σ_B	PPLTRST	PPLFAIR	PPLHLP	Σ_B	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	0.242	*0.754*	*0.369*	PPLTRST	0.684	*0.859*	*0.875*
PPLFAIR	0.144	0.151	*0.285*	PPLFAIR	0.613	0.744	*0.825*
PPLHLP	0.137	0.083	0.570	PPLHLP	0.580	0.571	0.643

(b) Household- or address-based sample

	The Netherlands				**Israel**		
Σ_T	PPLTRST	PPLFAIR	PPLHLP	Σ_T	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	4.181	*0.522*	*0.365*	PPLTRST	5.507	*0.546*	*0.470*
PPLFAIR	1.813	2.882	*0.381*	PPLFAIR	2.899	5.112	*0.480*
PPLHLP	1.394	1.207	3.489	PPLHLP	2.508	2.469	5.172
Σ_W	PPLTRST	PPLFAIR	PPLHLP	Σ_W	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	3.998	*0.521*	*0.360*	PPLTRST	5.108	*0.517*	*0.467*
PPLFAIR	1.750	2.817	*0.383*	PPLFAIR	2.473	4.487	*0.466*
PPLHLP	1.336	1.191	3.442	PPLHLP	2.248	2.102	4.540
Σ_B	PPLTRST	PPLFAIR	PPLHLP	Σ_B	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	0.191	*0.593*	*0.644*	PPLTRST	0.429	*0.861*	*0.581*
PPLFAIR	0.067	0.066	*0.296*	PPLFAIR	0.468	0.689	*0.631*
PPLHLP	0.061	0.017	0.047	PPLHLP	0.308	0.424	0.654
	Lithuania				**Greece**		
Σ_T	PPLTRST	PPLFAIR	PPLHLP	Σ_T	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	5.521	*0.573*	*0.4*	PPLTRST	5.479	*0.676*	*0.579*
PPLFAIR	3.012	5.008	*0.519*	PPLFAIR	3.452	4.757	*0.625*
PPLHLP	2.746	2.768	5.692	PPLHLP	2.975	2.994	4.821
Σ_W	PPLTRST	PPLFAIR	PPLHLP	Σ_W	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	4.389	*0.521*	*0.421*	PPLTRST	4.261	*0.614*	*0.531*
PPLFAIR	2.261	4.289	*0.468*	PPLFAIR	2.479	3.827	*0.575*
PPLHLP	1.873	2.059	4.513	PPLHLP	2.074	2.129	3.585
Σ_B	PPLTRST	PPLFAIR	PPLHLP	Σ_B	PPLTRST	PPLFAIR	PPLHLP
PPLTRST	1.003	*0.781*	*0.685*	PPLTRST	1.161	*0.911*	*0.728*
PPLFAIR	0.605	0.598	*0.711*	PPLFAIR	0.917	0.872	*0.797*
PPLHLP	0.683	0.547	0.990	PPLHLP	0.844	0.800	1.555

covariances on the between-interviewer level. If so, the factor solution based on the total (co)variances may be biased.

To answer this question, a stepwise procedure is carried out, incrementally restricting the parameters at the between-interviewer level. Table 5 shows these steps: groups of parameters are systematically omitted at the between-interviewer level. It can then be formally tested whether the factor solution at the between level can be ignored without decreasing the global model fit of the multilevel factor analysis. This model reduction process is performed in three consecutive steps.

Step zero is the starting point, where all the (co)variances are free in both the between and within matrices. First (step one), the relationships between all pairs of items belonging to different latent constructs are omitted in the between-interviewer matrix $\mathbf{\Sigma_B}$, as can be seen in table 5, obtaining 27 degrees of freedom.

Each item of the first construct has six relationships with the items of the two other constructs, and the three items of the second construct have three relationships with the items of the last construct. Next (step two), all relationships between items of the same construct are ignored in $\mathbf{\Sigma_B}$, obtaining nine additional degrees of freedom. In the last step (step three), the variances of the items are also omitted in the between-level factor solution, again obtaining nine additional degrees of freedom. After this step, the between-covariance information is completely ignored (or $\mathbf{\Sigma_B} = 0$), assuming no differences between the interviewers whatsoever. During all the steps of the model reduction process, the within-level factor solution will not be restricted at all. Table 6 provides the results of the model reduction process.

In the first step, all covariances between items of different concepts are set to zero, resulting in a lack of model fit, particularly for Belgium, Slovenia, Israel, Lithuania, and Greece, and, to a lesser extent, Spain. Only in Finland and the Netherlands does the reduced model seem to be acceptable. Apart from Finland and the Netherlands, this may indicate that interviewer-specific deviations for an item related to a particular concept are predictive of interviewer-specific deviations on an item related to another concept. This also suggests that correlations between different concepts may be biased because of interviewer effects. In this regard, consider the example of Slovenia or Lithuania shown in figures 2 and 3. The figures show the factor solution based on the total covariance structure (left) compared with the factor solution based on the within-covariance structure alone (right), assuming that interviewer effects are completely incorporated in $\mathbf{\Sigma_B}$ (as in step zero). It appears from the figures that the correlations between the concepts are somewhat greater based on the total covariances than on the within-interviewer covariances, suggesting a slight overestimation of the covariances in the total covariance structure. However, it is not completely inconceivable that the between-interviewer covariance might be negative, implying that the relationships in the total covariance matrix are smaller than the ones in the within-interviewer covariance matrix. In this regard, consider two variables that are negatively correlated and where interviewers have the tendency to systematically

prioritize one end of the scale for the two survey questions, suppressing the real negative correlation.

In the second step, the covariances between the items of the same construct are also omitted. For all eight countries, this leads to a strong indication of a lack of fit of the newly specified factor model, at least when compared with the model specified in step one. This suggests that items of the same construct are correlated at the interviewer level, possibly biasing the true correlations. The example of Slovenia and Lithuania in figures 2 and 3 indeed shows that the factor loadings relating the items to their respective concepts are smaller for the within-interviewer covariance structure than for the total covariance matrix. This might lead to the suggestion that factor loadings are more likely to be smaller after removing the interviewer effects. However, and again not inconceivably, if some items are set in the reverse order to another item of the same construct, interviewer effects might also mitigate the expected correlation between the two items. Further research on this topic may examine interviewer effects on a balanced set of survey items, where items are both positively and negatively formulated.

The third step of the analysis also omits the variances of the items at the between-interviewer level. Except for Finland, this decreases the model fit. This may reflect the findings that are already observed in table 2, where the intraclass correlations are presented with respect to the nine items for the eight countries. After this final step, the entire between-interviewer covariance matrix is set to zero.

Not only do the point estimates of the elements of the covariance matrix differ between the total and the within-interviewer matrices, but their respective standard errors may also be affected. When the estimates of the respective covariance matrices are provided using maximum likelihood, the standard errors of the estimates using the within-interviewer covariances are somewhat larger than their total covariance counterparts. When the clustering in the data is ignored (this is exactly what the total covariance matrix is doing), the uncertainty about parameters is usually underestimated; therefore, standard errors usually increase when the clustering is taken into account. As a result, when survey researchers apply structural equation models or measurement models and do not specify the clustering in the data, standard errors may be too small (apart from the uncertainty due to possible bias). Considering figures 2 and 3 again, it can be observed that the standard errors of the estimates following the within-interviewer covariance are indeed somewhat larger than in the equivalent factor model, informed by the total covariance matrix.

For example, in Lithuania, the variance of the factor loading of the first item of the latent variable "Social trust" is 0.018^2 for the model based on the total covariance structure, and 0.021^2 when based on the within-interviewer covariances (see figures 2 and 3). The variance inflation factor for this particular parameter is $0.021^2/0.018^2 = 1.36$. Considering all 21 pairs of variances of the parameter estimates of Lithuania, as shown in this figure, the average variance inflation factor is 1.35 when comparing the model based on the total

Table 5. Stepwise Elimination of the Between-Interviewer Covariances for a Measurement Model Containing Nine Items and Three Latent Concepts, to be Executed for Each Country Separately

	PPLTRST	PPLFAIR	PPLHLP	TRSTPRL	TRSTPLT	TRSTPRT	IMBGECO	IMUECLT	IMWBCNT
Step zero									
Between-covariance Σ_B									
PPLTRST	free								
PPLFAIR	free	free							
PPLHLP	free	free	free						
TRSTPRL	free	free	free	free					
TRSTPLT	free	free	free	free	free				
TRSTPRT	free	free	free	free	free	free			
IMBGECO	free	free	free	free	free	free	free		
IMUECLT	free	free	free	free	free	free	free	free	
IMWBCNT	free	free	free	free	free	free	free	free	free
Within-covariance Σ_W									
PPLTRST	free								
PPLFAIR	free	free							
PPLHLP	free	free	free						
TRSTPRL	free	free	free	free					
TRSTPLT	free	free	free	free	free				
TRSTPRT	free	free	free	free	free	free			
IMBGECO	free	free	free	free	free	free	free		
IMUECLT	free	free	free	free	free	free	free	free	
IMWBCNT	free	free	free	free	free	free	free	free	free

Step one									
Between-covariance $\boldsymbol{\Sigma}_B$									
PPLTRST	free								
PPLFAIR	free	free							
PPLHLP	free	free	free						
TRSTPRL	0	0	0	free					
TRSTPLT	0	0	0	free	free				
TRSTPRT	0	0	0	free	free	free			
IMBGECO	0	0	0	0	0	0	free		
IMUECLT	0	0	0	0	0	0	free	free	
IMWBCNT	0	0	0	0	0	0	free	free	free
Within-covariance $\boldsymbol{\Sigma}_W$									
PPLTRST	free								
PPLFAIR	free	free							
PPLHLP	free	free	free						
TRSTPRL	free	free	free	free					
TRSTPLT	free	free	free	free	free				
TRSTPRT	free	free	free	free	free	free			
IMBGECO	free	free	free	free	free	free	free		
IMUECLT	free	free	free	free	free	free	free	free	
IMWBCNT	free	free	free	free	free	free	free	free	free

Continued

Table 5. *Continued*

	PPLTRST	PPLFAIR	PPLHLP	TRSTPRL	TRSTPLT	TRSTPRT	IMBGECO	IMUECLT	IMWBCNT
Step two									
Between-covariance Σ_B									
PPLTRST	free								
PPLFAIR	0	free							
PPLHLP	0	0	free						
TRSTPRL	0	0	0	free					
TRSTPLT	0	0	0	0	free				
TRSTPRT	0	0	0	0	0	free			
IMBGECO	0	0	0	0	0	0	free		
IMUECLT	0	0	0	0	0	0	0	free	
IMWBCNT	0	0	0	0	0	0	0	0	free
Within-covariance Σ_W									
PPLTRST	free								
PPLFAIR	free	free							
PPLHLP	free	free	free						
TRSTPRL	free	free	free	free					
TRSTPLT	free	free	free	free	free				
TRSTPRT	free	free	free	free	free	free			
IMBGECO	free	free	free	free	free	free	free		
IMUECLT	free	free	free	free	free	free	free	free	
IMWBCNT	free	free	free	free	free	free	free	free	free

Step three									
Between-covariance $\boldsymbol{\Sigma}_B$									
PPLTRST	0								
PPLFAIR	0	0							
PPLHLP	0	0	0						
TRSTPRL	0	0	0	0					
TRSTPLT	0	0	0	0	0				
TRSTPRT	0	0	0	0	0	0			
IMBGECO	0	0	0	0	0	0	0		
IMUECLT	0	0	0	0	0	0	0	0	
IMWBCNT	0	0	0	0	0	0	0	0	0
Within-covariance $\boldsymbol{\Sigma}_W$									
PPLTRST	free								
PPLFAIR	free	free							
PPLHLP	free	free	free						
TRSTPRL	free	free	free	free					
TRSTPLT	free	free	free	free	free				
TRSTPRT	free	free	free	free	free	free			
IMBGECO	free	free	free	free	free	free	free		
IMUECLT	free	free	free	free	free	free	free	free	
IMWBCNT	free	free	free	free	free	free	free	free	free

Table 6. Effects of the Stepwise Reduction of Between-Interviewer Covariance Matrix on the Global Fit of the Multilevel Measurement Model, ESS-5

		Finland	Belgium	Spain	Slovenia
(a) Individual-based sample					
Step one: omitting covariances of different construct	χ^2	27.53	85.02	41.11	68.82
	Df	27	27	27	27
	$p(M_0 \rightarrow M_1)$	0.44	0.00	0.04	0.00
Step two: omitting covariances of different and same construct	χ^2	51.38	111.73	134.17	247.55
	df	36	36	36	36
	$p(M_0 \rightarrow M_2)$	0.05	0.00	0.00	0.00
	$p(M_1 \rightarrow M_2)$	0.00	0.00	0.00	0.00
Step three: omitting all variances and covariances	χ^2	61.83	266.02	501.41	461.35
	df	45	45	45	45
	$p(M_0 \rightarrow M_3)$	0.05	0.00	0.00	0.00
	$p(M_2 \rightarrow M_3)$	0.31	0.00	0.00	0.00
		The Netherlands	Israel	Lithuania	Greece
(b) Household- or address-based sample					
Step one: omitting covariances of different construct	χ^2	36.40	75.96	56.49	61.43
	df	27	27	27	27
	$p(M_0 \rightarrow M_1)$	0.11	0.00	0.00	0.00
Step two: omitting covariances of different and same construct	χ^2	62.44	393.67	389.26	710.04
	df	36	36	36	36
	$p(M_0 \rightarrow M_2)$	0.00	0.00	0.00	0.00
	$p(M_1 \rightarrow M_2)$	0.00	0.00	0.00	0.00
Step three: omitting all variances and covariances	χ^2	102.49	1105.49	996.01	2078.33
	df	45	45	45	45
	$p(M_0 \rightarrow M_3)$	0.00	0.00	0.00	0.00
	$p(M_2 \rightarrow M_3)$	0.00	0.00	0.00	0.00

The *p*-values evaluate the model reduction tests (e.g., $p(M_0 \rightarrow M_1)$ evaluates whether the model in step one is as good as in step zero).

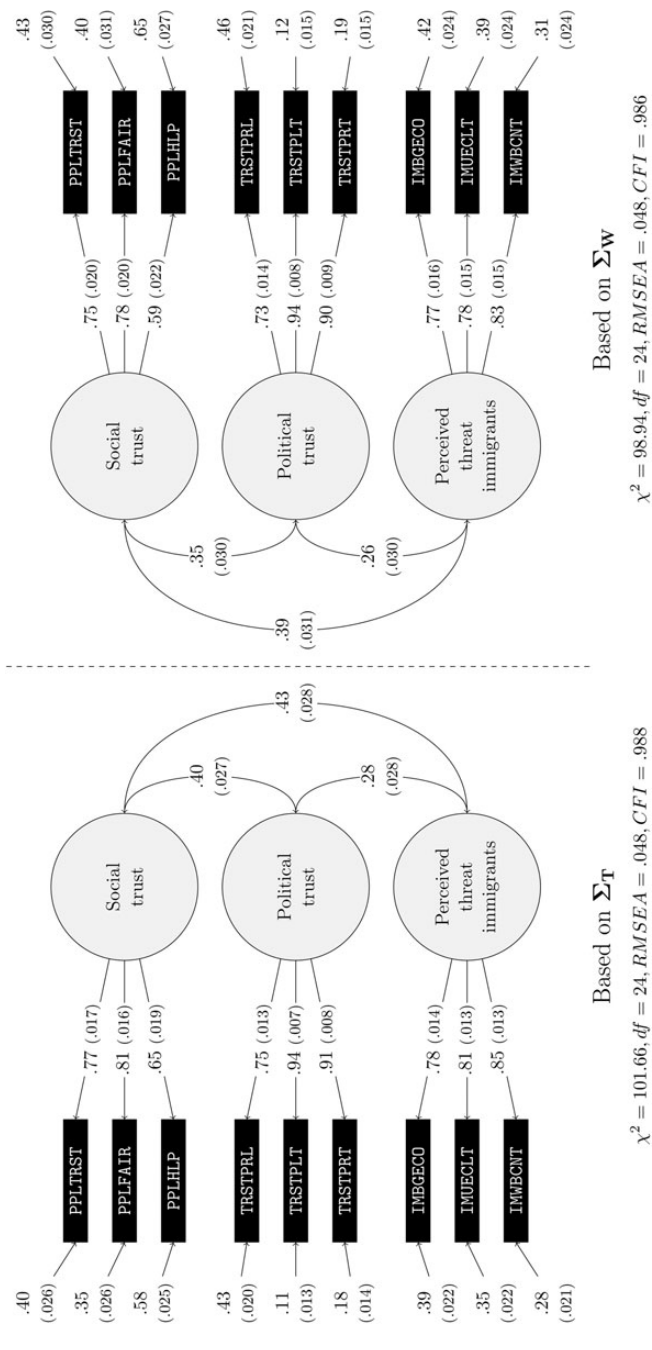

Figure 2. Measurement Models Based on Total $\boldsymbol{\Sigma}_{\mathbf{T}}$ and Within $\boldsymbol{\Sigma}_{\mathbf{w}}$ Covariance Structure, Standardized Estimates, Slovenia.

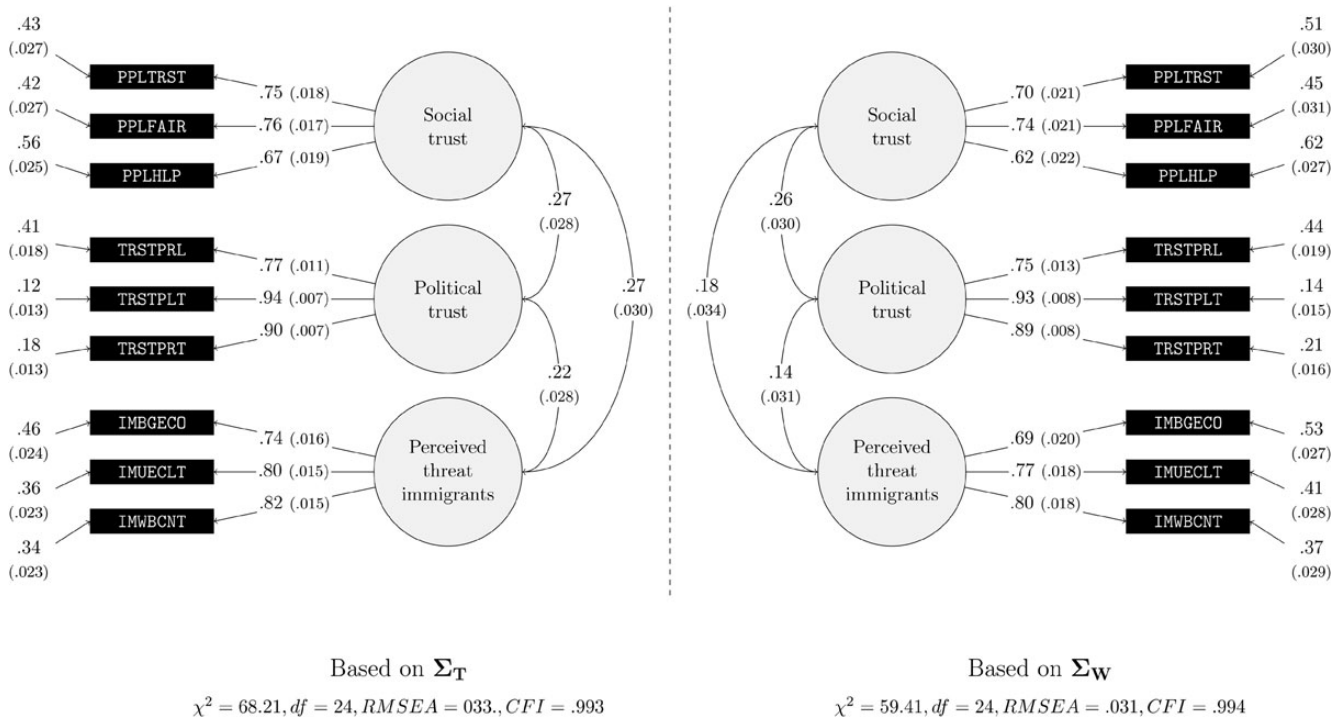

Figure 3. Measurement Models Based on Total $\mathbf{\Sigma_T}$ and Within $\mathbf{\Sigma_w}$ Covariance Structure, Standardized Estimates, Lithuania.

covariance structure to the one based on the within-interviewer covariance structure. In Finland, this variance inflation factor is 1.07 (though the standard errors are not shown).

An even more complicating circumstance is the fact that the standard errors strongly depend on the kind of estimator used in the analysis (Yuan and Hayashi 2005). Mplus software, which was used for the multilevel covariance structure analysis, offers a wide range of estimation methods, often leading to different standard errors. Given that maximum likelihood (ML) holds a position in the middle, robust maximum likelihood (MLR) provides much larger standard errors, whereas weighted least squares produce substantially smaller standard errors.

An example of an Mplus code to run a two-level confirmatory factor analysis is shown in the appendix.

4. DISCUSSION

The results of the analysis suggest that interviewer effects not only influence the variance of a single survey item, but also affect the relationships between survey variables. However, interviewer effects have more impact on the variances of survey items than on item covariances. Nonetheless, interviewer effects may still destabilize the estimates of a measurement model and additionally increase the standard errors of the estimates. This destabilizing effect does not necessarily mean that relationships between survey variables are systematically overestimated when interviewer effects are ignored. Although there were no clear examples found in the analyses covered by this paper, it is, in theory, possible that interviewer effects mitigate the strength of the relationships between survey variables, resulting in underestimated correlations.

As the ESS sample design in the various participating countries is not restricted by an interpenetration of interviewers and areas, these two factors are hard to disentangle. This means that the step from the total covariances to the within-interviewer covariances may take away too much information. Therefore, taking the within-interviewer covariance structure to an extent acts as a "worst-case scenario" benchmark. Nonetheless, as interviewer effects are widely recognized, simply ignoring the problem and solely relying on the total covariance matrix may also be too naive.

As both estimates may be somewhat biased and their standard errors are underestimated when interviewer effects are ignored in the measurement model, survey researchers must be aware of the fact that survey data cannot be used without acknowledging the data flaws due to interviewer effects. Therefore, it is advisable to use advanced software tools that accommodate for such interviewer effects. If these would not be available, researchers are urged to carefully adapt their expectations with regard to the power of the survey data and act in a more conservative or reserved way when testing hypotheses.

Although interviewers may still be seen only as facilitators of the survey process, both during the contact phase and during the actual interviewing, they are a substantial part of the measurement instrument and thus affect the data that are produced. This inevitably implies that the quality of the obtained survey data cannot be assessed without a minimal description and documentation of the interviewers. Supporting what Schnell and Kreuter (2005) propose, data sets should also contain a variable identifying the interviewer (and/or area identifiers), enabling researchers to take into account their effects during the analysis. Currently, the European Social Survey does not include these local area or sampling point identifiers in its main data sets, although separate files have been made available since round 5 containing this information.

Although more research would be preferable, it seems that countries that have to deal with considerable levels of inter-interviewer variance in the univariate sense also seem to face more interviewer effects with respect to relationships between survey variables. Therefore, interviewer variance on single items may be considered an indication for interviewer effects on factor analysis of structural equation models. In this regard, the possibility cannot be excluded that country differences with regard to interviewer (co)variances also affect the assessment of measurement equivalence in cross-national or cross-cultural survey research.

This paper extended research on interviewer effects from univariate or descriptive statistics to multivariate statistics. Nevertheless, as it seems that countries showing high levels of interviewer effects on descriptive statistics also show high levels of interviewer effects on multivariate statistics, a common cause may be searched for, trying to explore the underlying mechanisms leading to these interviewer effects. We therefore speculate that specific interviewer behavior that deviates from the norm of standardized interviewing may be an important determinant of interviewer effects. Such norms of standardized interviewing may include reading the exact words of the question, reading all items on the response scale as instructed by the questionnaire, neutrality, probing when the initial answer is not an available option, and so forth. This suggests that whenever interviewers provide their respondents with similar or standardized stimuli during the interview, fewer interviewer effects should be observed in the data. So-called paradata such as audiotapes of the interview may be valuable tools in this respect. Regarding the differences between countries, the European Social Survey is typically a survey that has a central coordinating team developing the questionnaires and prescribes the basic methodological standards, whereas participating countries and particularly the local fieldwork organizations are left with considerable leeway with regard to interviewer selection, training, and remuneration. It may therefore be worthwhile to closely monitor the production processes of the fieldwork for the different participating countries. Again, this requires fieldwork and organizational paradata in order to learn the best practices from countries showing low levels of interviewer effects.

Appendix

An example of Mplus code in order to specify a two-level confirmatory factor analysis is provided here. In this particular example, a model is specified where the covariance matrices at both the respondent and the interviewer level are restricted to only have free relationships between the items belonging to the same construct.

The input datafile is a respondents by variables datafile. The first command lines (TITLE, DATA, VARIABLES, CLUSTER, and ANALYSIS) are straightforward. However, the specifications of the model should be discussed in more detail. In a two-level covariance model, the model structure at the two different levels has to be specified separately. The %WITHIN%-statement specifies the factor structure on the respondent level, whereas the % BETWEEN%-statement specifies the factor structure on the interviewer level. At both levels, three latent constructs are specified, informed by three sets of observed items. This means that in this particular example, six latent variables will be constructed, three at the respondent level (soctr_w, poltr_w, and immig_w) and three at the interviewer level (soctr_b, poltr_b, and immig_b).

```
TITLE: Two-level CFA
DATA: File is file.dat;
VARIABLE:
NAMES are int ppltrst pplfair pplhlp trstprl trstplt trstprt imbgeco imueclt
imwbcnt;
USEVAR are int ppltrst pplfair pplhlp trstprl trstplt trstprt imbgeco imueclt
imwbcnt;
MISSING are int ppltrst pplfair pplhlp trstprl trstplt trstprt imbgeco imueclt
imwbcnt (999);
CLUSTER = INT;
ANALYSIS:
TYPE = twolevel;
ESTIMATOR=ML;
MODEL:
%WITHIN%
soctr_w by ppltrst pplfair pplhlp;
poltr_w by trstprl trstplt trstprt;
immig_w by imbgeco imueclt imwbcnt;
%BETWEEN%
soctr_b by ppltrst pplfair pplhlp ;
poltr_b by trstprl trstplt trstprt;
immig_b by imbgeco imueclt imwbcnt;
OUTPUT: sampstat standardized (stdyx);
```

References

Beullens, K., and G. Loosveldt (2013), *Assessing Interviewer Variance and Its Implication for Data Collection*, *Deliverable 12.10* (Tech. Rep.), KU Leuven, ESS-DACE.

Biemer, P., and L. Lyberg (2003), *Introduction to Survey Quality*, Hoboken, NJ: Wiley.

Davis, P., and A. Scott (1995), "The Effect of Interviewer Variance for Domain Comparisons," *Survey Methodology*, 21, 99–106.

Freeman, J., and E. Butler (1976), "Some Sources of Interviewer Variance in Surveys," *Public Opinion Quarterly*, 40(1), 79–91.

Groves, R., and L. Magilavy (1986), "Measuring and Explaining Interviewer Effects in Centralized Telephone Surveys," *Public Opinion Quarterly*, 50(2), 251–266.

Kish, L. (1962), "Studies of Interviewer Variance for Attitudinal Variables," *Journal of the American Statistical Association*, 57, 92–115.

Loosveldt, G. (2008), "Face-To-Face Interviews," in *International Handbook of Survey Methodology*, eds. de Leeuw, E., J. Hox, and D. Dillman, pp. 201–220, New York: Erlbaum.

Loosveldt, G., and K. Beullens (2010), *Evaluation of Interviewer Effects in Round Four of the European Social Survey*, *Deliverable 12.2* (Tech. Rep.), KU Leuven, ESS-DACE.

Mahalanobis, P. (1946), "Recent Experiments in Statistical Sampling in the Indian Statistical Institute," *Journal of the Royal Statistical Society*, 109, 325–370.

Mangione, T., F. Fowler, and T. Louis (1992), "Question Characteristics and Interviewer Effects," *Journal of Official Statistics*, 8(3), 293–307.

Muthén, B. (1989), "Latent Variable Modeling in Heterogeneous Populations," *Psychometrika*, 54 (4), 557–585.

Muthén, B. (1994), "Multilevel Covariance Structure Analysis," *Sociological Methods and Research*, 22(3), 376–398.

O'Muircheartaigh, C., and P. Campanelli (1998), "The Relative Impact of Interviewer Effects and Sample Design Effects on Survey Precision," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 161(1), 63–77.

Philippens, M., and G. Loosveldt (2004), "Interviewer-Related Variance in the European Social Survey," Paper presented at the Sixth International Conference in Logic and Methodology, Amsterdam, August.

Rao, J. (2005), "On Measuring the Quality of Survey Estimates," *International Statistical Review*, 73(2), 241–244.

Schnell, R., and F. Kreuter (2005), "Separating Interviewer and Sampling-Point Effects," *Journal of Official Statistics*, 21(3), 389–410.

Sedransk, J., and R. Tourangeau (2013), "A Statement of the Editors," *Journal of Survey Statistics and Methodology*, 1(1), 1–5.

West, B., F. Kreuter, and U. Jaenichen (2013), "Interviewer Effects in Face-to-Face Surveys: A Function of Sampling, Measurement Error or Nonresponse," *Journal of Official Statistics*, 29(2), 277–297.

West, B., and K. Olson (2010), "How Much of Interviewer Variance Is Really Nonresponse Error Variance?" *Public Opinion Quarterly*, 74(5), 1004–1026.

Yuan, K.-H., and K. Hayashi (2005), "On Muthén's Maximum Likelihood for Two-Level Covariance Structure Models," *Psychometrika*, 70(1), 147–167.