Survey Research Methods (2023)
Vol. 17, No. 1, pp. 51-62
doi:10.18148/srm/2023.v17i1.8011
European Survey Research Association

Religious involvement across Europe: Examining its measurement comparability

Alisa Remizova[1], Eldad Davidov[1,2], and Maksim Rudnev[3,4]
[1]University of Cologne
[2]University of Zurich
[3]University of Waterloo
[4]HSE University

Religion permeates many aspects of individuals' lives and interplays with a multitude of other theoretical constructs of interest for researchers. The cross-national European Social Survey (ESS) includes the Religious Involvement scale for the comparative study of religiosity. However, the cross-national comparability of this scale has been investigated only occasionally. The current study examined the measurement invariance of the Religious Involvement scale across ESS countries in nine survey rounds (2002–2018). Our findings demonstrated that exact measurement invariance across all countries was not supported by the data, however approximate invariance could be established using the alignment procedure. The frequency of praying item was more invariant than the frequency of religious attendance and self-assessed religiosity indicators. Finally, we listed the comparable estimates of a mean country religious involvement. We conclude that the Religious Involvement scale in the ESS can be confidently used for the comparative study of religious involvement relying on the means estimated with the alignment method.

Keywords: religious involvement; measurement invariance (MI); European Social Survey (ESS); multiple group confirmatory factor analysis (MGCFA); alignment; exact vs. approximate measurement invariance

1 Introduction

Religiosity is related to many social outcomes at the interpersonal and intergroup levels and has been employed by various researchers to examine characteristics of individuals and societies. With the rising number of political debates involving religion, religious heterogeneity, and migrant flows in European countries (Davie & Leustean, 2021; Triandafyllidou & Magazzini, 2021), empirical analyses of global processes with valid measures of religiosity are becoming increasingly relevant. Large-scale repeated surveys, which involve many countries, opened up possibilities for comparative research on religiosity across nations and time points. In particular, the European Social Survey (ESS) allows the analysis of religiosity in the majority of European populations as well as in a few countries outside this region. The core ESS questionnaire includes the three-item Religious Involvement (RI) scale (Meuleman & Billiet, 2011), which has been widely used across contexts to explain various psychosocial phenomena such as human values (Schwartz, 2012), life satisfaction (Joshanloo, 2016), prejudice against migrants (Bohman & Hjerm, 2014), political voting (Montgomery & Winter, 2015), or marriage attitudes (Liefbroer & Rijken, 2019) to name a few.

Despite many advantages of the comparative study of religiosity, it also poses a number of methodological challenges. Religiosity is one of the most complex constructs to measure due to the cross-country diversity in its semantics (Finke & Bader, 2017; Hill & Pargament, 2003; Tarakeshwar et al., 2003). The variations in social structures, policies, and cultural and religious norms across populations, which give multinational studies their value, can cause a different understanding of religiosity as well as varying interpretations of survey questions by respondents and, therefore, undermine their comparability. If these dissimilarities are present but not accounted for, then the cross-country differences in religiosity and its relations with other theoretical constructs may reflect not only true variations in individuals' characteristics but also measurement artifacts (Chen, 2008; Vandenberg & Lance, 2000). Meaningful comparisons require that the same construct is measured in the same way across all countries (Davidov et al., 2014; Jowell, 1998). To ensure

Contact information: Alisa Remizova, Faculty of Management, Economics and Social Sciences, Institute of Sociology and Social Psychology, University of Cologne (E-mail: remizova@wiso.uni-koeln.de).

it, researchers need to guarantee the invariance of their measurement instruments (hereafter measurement invariance is abbreviated as MI).

To date, only few studies have tested MI of the RI scale before utilizing it (e.g., Lemos et al., 2019; Meuleman & Billiet, 2011). Using the multiple group confirmatory factor analysis (MGCFA) approach, which requires the exact similarity of measurement parameters, the researchers were able to establish the exact invariance of the scale only for two religiousity indicators and only for a subset of European countries (Lemos et al., 2019; Meuleman & Billiet, 2011). These findings challenged the assumption that the RI can be used for carrying out a meaningful cross-national assessment of religiosity and its mean comparison.

In this study, we addressed the lack of invariance of the RI scale and focused on a more liberal MI test rather than on the exact one. We applied the alignment optimization procedure (Asparouhov & Muthén, 2014) based on the idea that the comparisons of latent means are unbiased if the parameter differences across units are small enough. We investigated whether religious involvement can still be validly compared across countries participating in the ESS even when exact MI is not given.

We begin by discussing the concept of MI. Next, we provide an overview of the RI scale and potential factors that may undermine its cross-national comparability. In the following section, we describe the data, measures, and analytical strategy underlying this study. The results of MI tests for nine ESS rounds are presented in the next section. Finally, we discuss the main findings and consider areas for further research.

2 Background

2.1 Measurement invariance in comparative research

Measurement invariance or equivalence implies that the nature and understanding of the construct and corresponding measurement questions are the same in different groups of respondents and are not a function of their group belongingness (Davidov et al., 2014). If this is not the case, the parameters associated with the instrument, such as its means or relations with other constructs, may not be meaningfully compared across groups.

In recent years, various approaches have been developed to assess the invariance of multi-item measures (Millsap, 2011). They examine whether observed indicators of a construct relate to the latent variable in the same way across all groups (Milfont & Fischer, 2010). Among these methods, MGCFA has been used the most (Davidov et al., 2014). Researchers typically assess three hierarchically ordered levels of MI, defined by the constraints on the indicator parameters (Vandenberg & Lance, 2000). The lowest level of invariance, configural, implies the same conceptualization of the latent variable across groups. This is indicated by a similarity of the general factor structure, that is, the number of factors, configurations of items, and pattern of loadings. If configural invariance is supported by data, it is possible to compare the directions of associations of the construct with other variables of interest.

Metric (or weak) invariance guarantees the same units of the factor(s) across groups. In other words, when respondents react to the items through selecting particular categories, they express an equivalent degree of corresponding characteristics, given the same level of the latent variable. This level of invariance requires factor loadings to be equal. If it is supported by the data, one can validly compare the correlates between the scale and other constructs.

In addition to the metric invariance, the third, scalar (or strong) level of invariance implies that the latent variable score has the same zero point across groups. It can be established by examining if the loadings and intercepts of indicators are equal. Along with the research possibilities that previous levels of invariance allow, the factor means can additionally be compared across groups.

It has recently been pointed out that the MGCFA approach may be too strict when analyzing real survey data with a large number of countries (Davidov et al., 2014; B. Muthén & Asparouhov, 2013; van de Schoot et al., 2013). In some cases, scalar invariance is rejected even when the measurement differences are not critical. Various scholars have proposed different strategies to deal with noninvariance. Thus, partial invariance can allow meaningful analyses to be performed even when a small subset of statistical parameters is not exactly equal (Byrne et al., 1989; Steenkamp & Baumgartner, 1998). However, some researchers have shown that the comparison of factor means can still be distorted by noninvariant items (e.g., De Beuckelaer & Swinnen, 2011) Another proposed strategy has been to delete noninvariant indicators or to drop countries from the sample (Byrne & van de Vijver, 2010). However, a significant amount of valuable information can be lost using this approach (Davidov et al., 2014).

Asparouhov and Muthén (2014) introduced a different method to deal with the absence of metric and scalar invariance. Instead of dropping items or countries, or releasing equality constraints for some indicators as in the case of partial invariance, they suggested using the alignment optimization procedure to find a model that is as invariant as possible. This procedure is able to validly estimate latent means when there is a relatively small variability in measurement parameters in the data. Instead of constraining parameters to be equal across groups, alignment uses a simplicity function that estimates them in such a way that their cross-group differences are minimized. The fit of the model is the same as in the configural model, and the means of the latent variable are calculated with the best possible comparability that can be achieved in the data. Alignment also provides a complete

report on the degree of noninvariance of every model parameter in every group. In the next sections, we present the religiosity measures we studied and examine their invariance properties using the alignment procedure.

2.2 Comparability of religiosity measures in cross-country surveys

Measures of religiosity are included in many large-scale multinational surveys. As most surveys are not focused on the religious sphere only, they usually contain a rather limited number of the corresponding questions, which are administered to all studied populations. Such indicators have been criticized, among other things, for being too broad and insensitive to the specifics of religious expressions across contexts (Hill & Pargament, 2003, 2017). Cross-national surveys have been coordinated predominantly by Western scholars (Norris & Inglehart, 2004). As a consequence, the questions tend to be designed from a Western perspective and validated within Protestant or Catholic countries. The comparability of such measures may be undermined when multiple religions and cultures are included in the sample (Finke & Bader, 2017). After all, different religions may entail dissimilar beliefs and practices. Moreover, the differences can also refer to variations even within traditions and can depend on the specific historical, cultural, or political aspects of countries (Hill & Pargament, 2003; Saroglou & Cohen, 2013; Tarakeshwar et al., 2003). As a result, even the seemingly universal forms of worship and other religious expressions may not generalize across all populations (see, e.g., Bechert, 2018; Finke & Bader, 2017; Lemos et al., 2019; Meuleman & Billiet, 2011; Remizova et al., 2022; Siegers, 2011; Smith, 2017; Wilson, 1998). This implies that the cross-national MI of religiosity can be challenged.

A few studies, however, have examined the invariance properties of religiosity measures included in multinational surveys (Bechert, 2018; Lemos et al., 2019; Meuleman & Billiet, 2011; Remizova et al., 2022; Siegers, 2011). In particular, Meuleman and Billiet (2011) focused on public and private worship activities and religious self-assessment of European respondents, indicators they postulated to form a scale measuring a single religious involvement factor (see Figure 1).

Using MGCFA, Meuleman and Billiet (2011) provided evidence that the scale met the criteria of only partial metric invariance for the whole European sample of the second ESS round and partial scalar invariance for 21 of the 25 countries. Compared to the frequency of praying and self-assessed religiosity indicators, the frequency of attendance of religious services item differed across more European countries. Further evaluation of invariance properties of religiosity indicators across Christian populations also confirmed the attendance of religious services item being more problematic for cross-national comparisons than other two variables (Lemos et al., 2019). Among noninvariant countries, RI in Turkey was constantly displayed as not fully equivalent to that in Europe (Meuleman & Billiet, 2011). Thus, noninvariance was found for the frequency of praying indicator, with the intercept being higher than this parameter in invariant countries. Moreover, the attendance of religious services was hardly related to the construct in Muslim Turkey. According to religious norms of Islam and in contrast to Christianity, the area of public worship services is a men's domain. While men are expected to attend Friday prayers in the mosque, Muslim women can fulfill this obligation individually, pray at home, and do not participate in mosque worship (McAndrew & Voas, 2011; Pew Research Center, 2016). As a result, the attendance of religious services may be an irrelevant indicator when one compares the religiosity of Muslim and Christian females due to the gender-specific norms of religious practices.

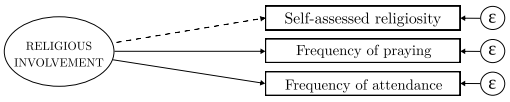


Figure 1

Measurement model of Religious Involvement scale

In another study, which involved European countries and was based on a discrete concept of religiosity, the frequency of religious attendance item was noninvariant across Protestant and Catholic countries (Siegers, 2011). It was explained by the fact that, compared to Catholics, Protestants deemphasized the role of the church in maintaining their religiosity and prioritized personal experiences of transcendence for the maintenance of faith.

In sum, previous evidence suggests that the meaning of religiosity may differ across European countries. However, does it vary to such an extent that deems the entire RI scale to be invalid for comparisons across these nations? To address this question, we investigated whether the RI measurement can be considered as similar across ESS countries in a way that allows a meaningful assessment of cross-national differences in religiosity. Extending Meuleman and Billiet (2011) study, we examined the exact as well as approximate invariance of the RI scale in all ESS rounds to date.

3 Data and methods

3.1 Data

For the empirical analysis, we employed ESS data from all nine currently available rounds collected in 2002-3, 2004-5,

2006-7, 2008-9, 2010-11, 2012-13, 2014-15, 2016-17, and 2018-19 (European Social Survey Rounds 1-9 Data, 2002-2021). All samples are representative of the population aged 15 and above in each country. The total number of observations across rounds was 451,175. As the share of missing individual responses was only 3%, we deleted these cases listwise (Tabachnick & Fidell, 2012). This removal did not largely change the means and standard deviations of the variables further used in the analysis.[1] The resulting sample included 437,990 respondents. Table A1 in the Appendix presents the final sample size per country and for each ESS round. Further information about the documentation of data collection and the data files are provided on the ESS website (European Social Survey, 2022).

3.2 Measures

Our analysis included three items including an individual's self-assessment of religiosity and two religious practices, namely, praying and service attendance (Meuleman & Billiet, 2011). The self-assessed religiosity was measured with the question "Regardless of whether you belong to a particular religion, how religious would you say you are?", with response options ranging from 0 (not at all) to 10 (very religious). The frequency of attendance of religious services and the frequency of praying were each assessed with the following questions: "Apart from special occasions such as weddings and funerals, about how often do you attend religious services currently?" and "Apart from when you are at religious services, how often, if at all, do you pray?", respectively. Response options for both items ranged from 1 (every day) to 7 (never). They were recoded in such a way that higher scores implied higher religious involvement.

3.3 Analytical procedure

We first tested the exact measurement invariance across countries in each round using the MGCFA approach. The nonnormality of religious indicators was accounted for by robust maximum likelihood (MLR) estimator (Kaplan et al., 2009)[2]. We followed a bottom-up strategy and started with the least restrictive invariance model—configural. The loading of the self-assessed religiosity indicator selected as a marker variable was fixed to 1 for identification purposes (Johnson et al., 2009). Next, we consecutively introduced equality constraints on the factor loadings of indicators for the metric model and on the factor loadings and intercepts for the scalar model. We evaluated the absolute fit of the models and the differences in fit statistics between levels of invariance with three measures: the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean square residual (SRMR). A CFI larger than 0.90, and an RMSEA and SRMR smaller than 0.08 were used as indications of satisfactory absolute fit (Browne & Cudeck, 1992; Hu & Bentler, 1999; Marsh et al., 2004). A change smaller than 0.01 in CFI, 0.015 in RMSEA, and 0.03 in SRMR between configural and metric models implied that metric invariance was achieved (Chen, 2007). The changes smaller than 0.01 in CFI, 0.015 in RMSEA, and 0.01 in SRMR between metric and scalar models were considered as acceptable and supportive of scalar invariance. The difference test based on chi-square values was not used in the study as it is overly strict with large sample sizes (Yuan & Chan, 2016). MGCFA analysis was carried out with the R program and the *lavaan* package (Rosseel, 2012).

We then turned to the alignment optimization procedure (Asparouhov & Muthén, 2014). Although the approximate invariance could be investigated with other methods such as the Bayesian approach (B. Muthén & Asparouhov, 2013) as well, we focused on alignment because it provides information about the degree of approximate invariance of each parameter and in each group, making it possible to detect noninvariant parameters in a convenient way.

The alignment was run with the MLR estimator as well. Initially, the model with the factor means and variances fixed to 0 and 1, respectively, was estimated, and no constraints were placed on the loadings and intercepts of indicators. Then, the factor means and variances were calculated with a simplicity function, which identifies the optimal measurement invariance pattern in a model. We initially employed the free alignment with the means freely estimated in all groups and then turned to the fixed alignment, which fixes the mean to 0 in a selected group, if necessary for a better model identification.

We used two indices to assess the degree of invariance of each measurement parameter: the R^2, which ranges from 0 to 1, with higher values indicating higher invariance (values of 0.90 were considered rather high, see e.g., Flake and McCoach, 2018, for the ordinal case); and a fit function contribution, whose higher value, in contrast, suggested that a parameter was more likely to be noninvariant compared to the other parameters of the same model (B. Muthén & Asparouhov, 2018). We also inspected the percentage of invariant and noninvariant countries across indicators. Asparouhov and Muthén (2014) suggested that the share of noninvariant parameters in a model may be as high as 20% to be able to rely on the latent mean estimates, or, as further stud-

[1] The listwise deletion only slightly increased the standard deviation for the self-assessed religiosity and frequency of religious attendance indicators, with a difference of 0.01 for both indicators on the full sample on nine survey rounds.

[2] Although the weighted least squares (WLS) estimator was employed in the original study (Meuleman & Billiet, 2011), we turned to the MLR estimation of the MGCFA models to ensure consistency with the further alignment analysis. At the moment the research was conducted, Mplus did not allow using WLS for the alignment. The skewness of indicators across rounds ranged from 0.61 to 0.83 for frequency of religious attendance, from 0.30 to 0.62 for frequency of praying, and from −0.16 to 0.03 for self-assessed religiosity.

ies demonstrated, raised to 25% and even higher under certain conditions (Flake & McCoach, 2018; B. Muthén & Asparouhov, 2014). We considered a finding of 25% or less noninvariant parameters as supportive of invariance. In addition, we conducted a Monte Carlo simulation study to evaluate the trustworthiness of aligned factor means. To do so, we generated the new artificial data using the final parameter estimates of the alignment models and then compared the generated estimates with our aligned estimates. We defined five hypothetical samples, with 100, 500, 1,000, 1,500, and 2,000 cases per group, including the condition corresponding to the samples in the ESS countries. For each condition, 500 replications were completed. A correlation of 0.98 or higher between the generated and estimated latent means was considered as evidence of high reliability of alignment results. Therefore, the aligned means could be compared across countries even if there is a large share of noninvariant parameters. All alignment and simulation analyses were conducted using the software package Mplus (L. K. Muthén & Muthén, 1998–2019). The R codes and Mplus files are available in the replication materials.

4 Results

Table 1 presents the results of the exact invariance tests of the RI scale for each ESS round separately. Global fit statistics for the configural models could not be assessed because the measurement consisted of only three indicators and so the model was just-identified [3]. Neither full metric nor full scalar invariance was supported by the data in all survey rounds, as indicated by the low fit of the models and the large changes in CFI, RMSEA, and SRMR for the scalar models. In the next step, we turned to the alignment procedure.

First, we ran a free alignment procedure. Based on its output, we employed the fixed optimization and specified the country with a factor mean value closest to 0 as the reference group. Table A2 in the Appendix presents the detailed alignment results for the factor loadings and intercepts of indicators. The percentage of noninvariant parameters exceeded the recommended 25% threshold in all surveys. Thus, the share of noninvariant countries averaged across factor loadings and intercepts of indicators was the highest in round 6, 51%, and the lowest in round 4, 42%. The mean R^2 across six parameters varied between 0.68 in round 7 with 22 countries and 0.82 in round 4 with 31 countries.

To compare the invariance of factor loadings and intercepts of the three indicators, we used the R^2 value and the share of noninvariant countries averaged across all survey rounds. We also examined the deviation of the fit function contribution from its mean value in a survey, as this index is survey-specific. The higher invariance of a parameter was indicated by the fit function contribution lower than the mean value. The results are presented in Table 2.

The frequency of praying item appeared to be the most invariant compared to the other two indicators. For both factor loading and intercept, the item had the highest R^2 and the lowest fit function contribution among parameters of the remaining variables as well as the lowest share of noninvariant countries among factor loadings. The lowest mean R^2 was evidenced for the self-assessed religiosity intercept and loading. In the majority of cases, this item contributed the most to the fit function with respect to noninvariant intercepts, while the frequency of religious attendance indicator contributed the most with respect to noninvariant loadings. The loading and intercept of the religious attendance item were noninvariant in a larger number of countries compared to parameters of the other items. In sum, this evidence suggested a higher noninvariance of the two indicators: self-assessed religiosity and frequency of religious attendance.

In some cases, noninvariance patterns could be attributed to the differences in the historical religious background of countries ((Inglehart & Welzel, 2005, 2020), see Table A1 in the Appendix for the classification of countries). For example, the loading of the religious attendance item was invariant in more than 50% of the historically predominant Protestant European countries (e.g., Norway) in every ESS round, but it was noninvariant in the majority of historically predominant Catholic (e.g., Belgium), Orthodox (e.g., Ukraine), and Islamic (e.g., Turkey) countries as well as in Israel (see Table A2 in the Appendix). The frequency of praying loading displayed invariance across the majority of historically Catholic, Protestant, and three Islamic countries included in the datasets.

The abovementioned results of the alignment suggest a low degree of invariance of the religious involvement measure. To evaluate whether the latent means could nevertheless be reliably recovered by alignment, we conducted Monte Carlo simulations. They were performed for five group sample sizes ranging between 100 and 2,000. The correlations of estimated and replicated means for the five conditions are presented in Table 3. Simulations with N = 100 indicated that religious involvement should not be compared when small samples are evaluated, as the correlations did not exceed the recommended threshold of 0.98. However, when the number of observations were similar to the actual ESS numbers (1,000–2,000, or even lower, i.e., 500), all associations were strong enough, 0.99 and higher, implying that the factor means are trustworthy[4]. In other words, the aligned

[3] We checked the standardized factor loadings for each country. They exceeded 0.60 for all items except the self-assessed religiosity in Albania in one survey round and the frequency of religious attendance in 5 of 238 cases across rounds—Iceland, Romania, Albania, and Turkey in two surveys. The factor loadings were above the threshold of 0.30 (Brown, 2015) in these countries, but not in Turkey with the value of 0.25 for the frequency of religious attendance in ESS 2. This specific Turkish sample in round 2 was thus excluded from further analyses

[4] There were some estimation issues for factor means, variances,

Table 1

MGCFA results: fit measures for metric and scalar measurement invariance

Model	CFI[a]	Δ CFI	RMSEA[b]	Δ RMSEA	SRMR[c]	Δ SRMR	χ^2	DF
ESS 1								
Metric	0.979	-	0.090	-	0.040	-	677.72	42
Scalar	0.843	0.136	0.173	0.084	0.083	0.043	4844.40	84
ESS 2								
Metric	0.977	-	0.096	-	0.042	-	865.44	48
Scalar	0.844	0.133	0.178	0.082	0.084	0.043	5721.14	96
ESS 3								
Metric	0.979	-	0.091	-	0.040	-	770.11	48
Scalar	0.867	0.111	0.160	0.069	0.078	0.038	4560.73	96
ESS 4								
Metric	0.976	-	0.095	-	0.042	-	1083.32	60
Scalar	0.872	0.104	0.156	0.061	0.078	0.035	5625.37	120
ESS 5								
Metric	0.978	-	0.094	-	0.042	-	951.21	54
Scalar	0.868	0.110	0.163	0.069	0.078	0.037	5544.24	108
ESS 6								
Metric	0.969	-	0.113	-	0.051	-	1357.18	56
Scalar	0.816	0.153	0.194	0.082	0.095	0.044	7851.63	112
ESS 7								
Metric	0.976	-	0.098	-	0.046	-	810.87	42
Scalar	0.854	0.123	0.173	0.075	0.084	0.038	4849.16	84
ESS 8								
Metric	0.977	-	0.099	-	0.046	-	853.74	44
Scalar	0.854	0.122	0.175	0.076	0.083	0.037	5135.62	88
ESS 9								
Metric	0.982	-	0.083	-	0.038	-	785.78	62
Scalar	0.869	0.113	0.160	0.076	0.077	0.039	5428.47	124

[a] comparative fit index, [b] root mean square error of approximation, [c] standardized root mean residual

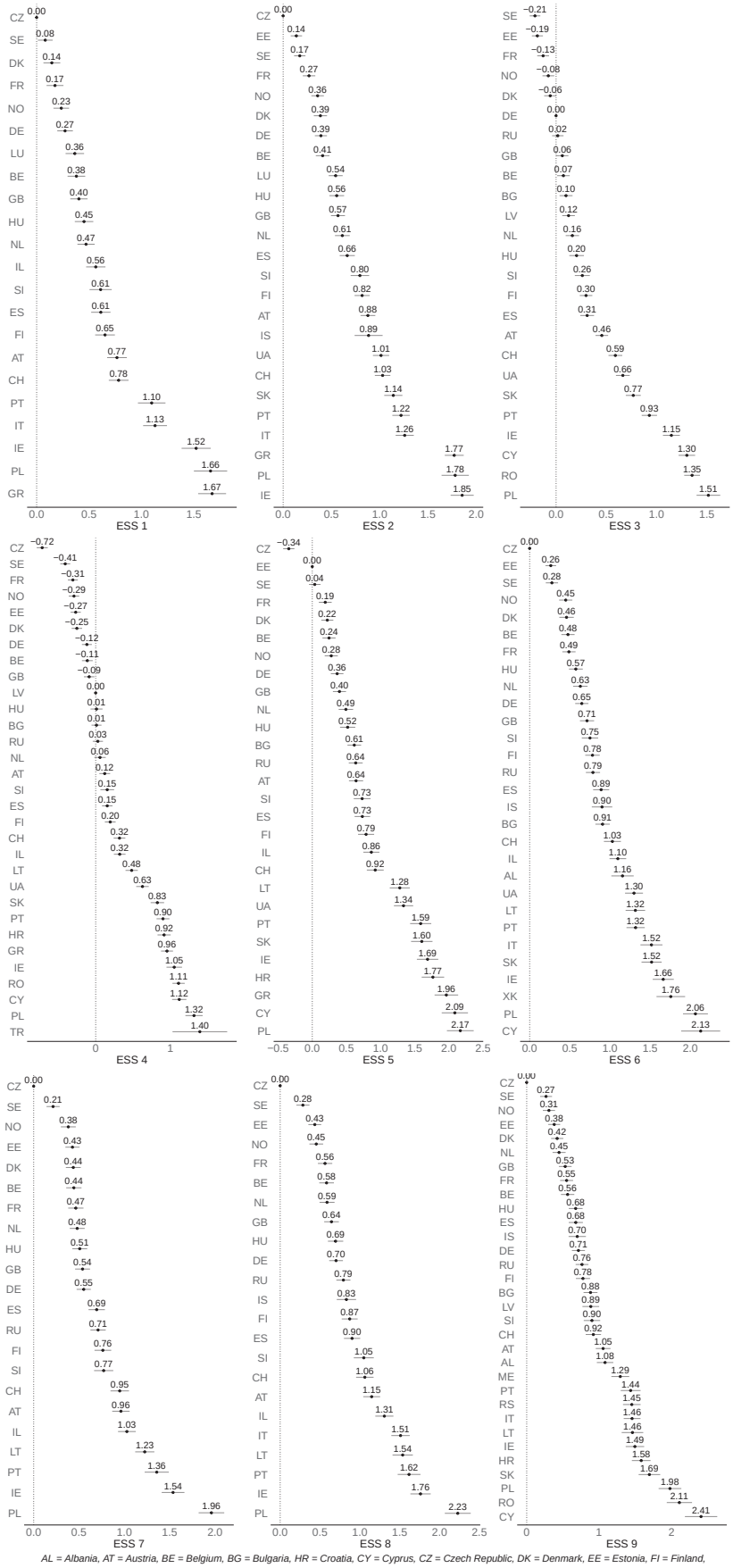
means may be compared across countries with confidence. These means are presented in Figure 2. They are the relative means of religious involvement in nine ESS rounds, with the means restricted to 0 in reference groups and other values estimated respectively to these groups. The correlations of means across surveys exceeded 0.92.

We also compared the factor means estimated by alignment and those by MGCFA under the assumption of partial scalar invariance for the second ESS round (as reported in Meuleman and Billiet, 2011). The means obtained with the two different methods correlated at 0.93 (N countries = 21). The most dissimilar means were found for Spain, Estonia, and Switzerland.

5 Discussion

The RI scale is used by many comparative researchers to examine religiosity or its relations with other phenomena of interest. However, the instrument has rarely been subjected to MI testing in previous studies. At the same time, invariance is an important requirement for cross-national research, as the lack of it may result in biased conclusions about similarities and differences across countries. In the current paper, we examined MI of the RI scale in nine rounds of the ESS. The previous study established partial metric invariance of the instrument for all countries and partial scalar invariance for a subset of nations using MGCFA and the data from the

and residual variances of indicators in simulations for ESS 2 and ESS 4. However, these problems were few and, in sum, came up only for five out of 3,060 parameters.



AL = Albania, AT = Austria, BE = Belgium, BG = Bulgaria, HR = Croatia, CY = Cyprus, CZ = Czech Republic, DK = Denmark, EE = Estonia, FI = Finland, FR = France, DE = Germany, GR = Greece, HU = Hungary, IS = Iceland, IE = Ireland, IL = Israel, IT = Italy, XK = Kosovo, LV = Latvia, LT = Lithuania, LU = Luxembourg, ME = Montenegro, NL = Netherlands, NO = Norway, PL = Poland, PT = Portugal, RO = Romania, RU = Russian Federation, RS = Serbia, SK = Slovakia, SI = Slovenia, ES = Spain, SE = Sweden, CH = Switzerland, TR = Turkey, UA = Ukraine, GB = United Kingdom

Figure 2

Means of religious involvement for the nine ESS rounds using alignment

Table 2

Alignment results: fit indices for intercepts and loadings of indicators

			Fit Function Contribution								
Parameter	NI %[a]	R^2	ESS 1	ESS 2	ESS 3	ESS 4	ESS 5	ESS 6	ESS 7	ESS 8	ESS 9
Frequency of religious attendance											
Intercept	0.62	0.90	9.0	9.3	7.9	33.4	13.5	15.1	3.4	5.8	6.7
Loading	0.46	0.55	4.0	4.6	−1.4	9.5	0.4	7.7	7.8	11.4	4.4
Self-assessed religiosity											
Intercept	0.53	0.81	16.6	19.0	22.4	12.8	23.2	41.2	12.6	15.5	44.1
Loading	0.41	0.34	−7.9	−7.4	2.0	−13.7	−3.6	−11.5	−8.9	−9.7	19.3
Frequency of praying											
Intercept	0.55	0.97	−7.9	−9.4	−10.5	−5.6	−12.7	−19.3	−2.0	−6.7	−2.1
Loading	0.19	0.87	−13.7	−16.0	−20.4	−36.5	−20.8	−33.2	−12.9	−16.4	−33.9

R^2 and the share of noninvariant countries are the averages across nine survey rounds. The fit contribution is presented separately for each survey round because it is not standardized; thus, its averaging across surveys would be biased. The positive values indicate the contribution higher than its average value in a survey, while the negative values indicate the contribution lower than its average value in a survey. Baseline group and number of groups are as follows: ESS 1: Czech Republic, 22; ESS 2: Czech Republic, 25; ESS 3: Germany, 25. ESS 4: Latvia, 31; ESS 5: Estonia, 28; ESS 6: Czech Republic, 29; ESS 7: Czech Republic, 22; ESS 8: Czech Republic, 23; ESS 9: Czech Republic, 32.

[a] Share of noninvariant parameters

Table 3

Monte Carlo simulations results of alignment: Correlations of estimated and replicated factor means averaged across 500 simulation runs

	Number of observations per group				
Round	100	500	1000	1500	2000
ESS 1	0.974	0.994	0.997	0.998	0.998
ESS 2	0.974	0.994	0.995	0.998	0.996
ESS 3	0.976	0.995	0.998	0.998	0.999
ESS 4	0.971	0.991	0.995	0.997	0.998
ESS 5	0.976	0.995	0.998	0.998	0.999
ESS 6	0.963	0.993	0.996	0.998	0.998
ESS 7	0.964	0.992	0.996	0.997	0.998
ESS 8	0.964	0.993	0.996	0.997	0.998
ESS 9	0.968	0.993	0.997	0.998	0.998

second ESS round (Meuleman & Billiet, 2011). These results did not allow the researchers to meaningfully compare the factor means across all countries in the sample. Along with this, it has been suggested that MGCFA is too strict for the analysis of many groups (Davidov et al., 2014; B. Muthén & Asparouhov, 2013; van de Schoot et al., 2013). We addressed this issue and employed the more lenient alignment, which tests approximate rather than exact invariance of measures and may allow factor means across countries to be validly assessed.

The findings of the MGCFA invariance tests revealed that the RI scale was not invariant across all countries in each ESS round. The alignment procedure indicated many noninvariant parameters but demonstrated ability to estimate reliable factor means. Therefore, the aligned means of religious involvement may be reliably compared despite the lack of exact invariance in some countries. Moreover, alignment indicated that the parameters of the frequency of praying item

were more invariant than the frequency of religious attendance and self-assessed religiosity indicators in almost all survey rounds. These results were in line with previous assessments of invariance properties of different religiosity indicators across countries worldwide, with far more heterogeneous samples compared to those of the ESS (Remizova et al., 2022). The frequency of religious attendance and self-assessed religiosity should be carefully used and cautiously interpreted in future studies when comparing religiosity patterns even across only (relatively homogenous) European populations. A scale with more items that demonstrated a relatively higher level of invariance would be better suited for a reliable cross-national survey measurement of religiosity. For example, such indicators could be the importance of religion in individuals' life or their confidence in religious institutions as included in the World Values Survey (Remizova et al., 2022).

The alignment also demonstrated cultural differences in the functioning of indicators across countries (Inglehart & Welzel, 2005, 2020). Thus, in line with Siegers (2011) findings, the frequency of religious attendance loading was invariant in the majority of Protestant countries in all rounds, but was not relevant for their comparison with Catholic populations as well as other societies (i.e., historically Islamic and Orthodox European countries and Israel) even after being aligned. Moreover, corresponding to Meuleman and Billiet (2011) results, we revealed that the frequency of praying loading was approximately invariant across most Catholic, Protestant, and three Islamic countries included in the samples.

Our results indicated a high correspondence between factor means estimated with alignment and those estimated with partial scalar invariance (Meuleman & Billiet, 2011), but there were dissimilarities for a few countries. Although the previous studies demonstrated the mixed evidence in regard to what method provides more reliable results (Marsh et al., 2018; Pokropek et al., 2019), the differences between the two methods found in the current study were not large suggesting that our results did not depend on the selected approach for the estimation of means of religious involvement across ESS populations. Based on the aligned means averaged across rounds, we conclude that the countries with the highest level of religious involvement are Poland, Cyprus, Kosovo, Greece, Romania, Ireland, Serbia, Croatia, Turkey, and Italy, while Czech Republic, Sweden, Estonia, Denmark, Norway, France, Belgium, Latvia, Germany, and the United Kingdom are the least religious.

This study has several limitations. First, the ESS includes only three indicators of religiosity, and they were mostly administered in European countries. Thus, our analysis had to focus on these indicators and on the religiosity dimension they measure in specific European societies. Other studies that include a richer set of items measuring religiosity also in countries outside of Europe may identify different cultural patterns of invariance. Indeed, the measurement of religiosity in the ESS might be problematic for a valid assessment of the concept, not only due to the low number of indicators it includes and its insufficient coverage of religiosity expressions, but also due to the fact that the use of only a few indicators requires to model the latent variable of religiosity as a unidimensional concept. This may result in mixing up different dimensions of religiosity in a single scale. However, previous research demonstrated the usefulness of unidimensional scales for measuring religiosity across countries that differ in their religious and cultural characteristics (Remizova et al., 2022).

Second, our research treated countries as homogeneous units, disregarding their regional and cultural variations. Future studies may examine the invariance properties of religiosity measures in a wider variety of settings and explore the differences within countries or across religious denominations.

Third, individuals' religiosity is not a consistent system across time as provided by a given religion (McGuire, 2008; Woodberry et al., 2020). The ESS has been collecting data for almost two decades, and it has been using the same items in each round. While this practice has the advantage of guaranteeing continuity in the measurement, it bears the risk that religious involvement could have undergone changes in its meaning in some societies that are not reflected in the scale. Thus, it could well be the case that indicators that operated well two decades ago may not do so anymore and should be reconsidered, especially when the number of countries increases, their diversity grows, or the general understanding of what it means to be a religious person changes.

Fourth, while the cultural zones framework allowed us to describe some differences in items functioning across countries, we did not systematically explain these variations. Future studies may assess, in a theoretically driven way, possible sources of noninvariance of religiosity measures by applying, for example, a multilevel structural equation modeling method that includes contextual variables in the analysis. Such an approach will make it possible to account for more specific differences across populations, such as their current religious composition, and the influence of these differences on patterns of noninvariance. Such research would deepen the understanding of cross-national noninvariance as well as ways to improve the survey assessment of religiosity.

Despite these limitations, our study is, to the best of our knowledge, one of the most extensive attempts to systematically examine the invariance properties of religiosity measurement across European countries. In spite of the differences in indicator parameters, using the alignment procedure we suggest that aligned means of religious involvement as measured in the ESS may be compared across countries in all survey rounds.

Acknowledgement

Eldad Davidov would like to thank the University of Zurich Research Priority Program "Social Networks" for their support during work on this study. The work of Maksim Rudnev is an output of a research project implemented as part of the Basic Research Program at the National Research University Higher School of Economics (HSE University). Maksim Rudnev's current affiliation is University of Waterloo. The authors would like to thank Lisa Trierweiler for the English proof of the manuscript

References

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(4), 495–508. https://doi.org/10.1080/10705511.2014.919210

Bechert, I. (2018). Comparing religiosity cross-nationally. *Zeitschrift für Religion, Gesellschaft und Politik*, *2*(1), 135–157. https://doi.org/10.1007/s41682-018-0016-z

Bohman, A., & Hjerm, M. (2014). How the religious context affects the relationship between religiosity and attitudes towards immigration. *Ethnic and Racial Studies*, *37*(6), 937–957. https://doi.org/10.1080/01419870.2012.748210

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford Press.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–466. https://doi.org/10.1037/0033-2909.105.3.456

Byrne, B. M., & van de Vijver, F. J. R. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, *10*(2), 107–132. https://doi.org/10.1080/15305051003637306

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*(5), 1005–1018. https://doi.org/10.1037/a0013193

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*(1), 55–75. https://doi.org/10.1146/annurev-soc-071913-043137

Davie, G., & Leustean, L. N. (2021). *The Oxford Handbook of Religion*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198834267.001.0001

De Beuckelaer, A., & Swinnen, G. (2011). Biased latent variable mean comparisons due to measurement non-invariance: A simulation study. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 117–147). Routledge.

European Social Survey. (2022). Data and Documentation. https://www.europeansocialsurvey.org/data/

European Social Survey Rounds 1-9 Data. (2002-2021). NSD - Norwegian Centre for Research Data, Norway Data Archive and distributor of ESS data for ESS ERIC.

Finke, R., & Bader, C. D. (2017). *Faithful measures: New methods in the measurement of religion*. New York University Press. https://doi.org/10.18574/nyu/9781479875214.001.0001

Flake, J. K., & McCoach, D. B. (2018). An investigation of the alignment method with polytomous indicators under conditions of partial measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(1), 56–70. https://doi.org/10.1080/10705511.2017.1374187

Hill, P. C., & Pargament, K. I. (2003). Advances in the conceptualization and measurement of religion and spirituality: Implications for physical and mental health research. *American Psychologist*, *58*(1), 64–74. https://doi.org/10.1037/0003-066x.58.1.64

Hill, P. C., & Pargament, K. I. (2017). Measurement tools and issues in the psychology of religion and spirituality. In R. Finke & C. D. Bader (Eds.), *Faithful measures: New methods in the measurement of religion* (pp. 48–77). NYU. https://doi.org/10.18574/nyu/9781479875214.001.0001

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Inglehart, R., & Welzel, C. (2005). *Modernization, cultural change and democracy: The human development sequence*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790881

Inglehart, R., & Welzel, C. (2020). The Inglehart-Welzel World Cultural Map. World Values Survey 7 [Provisional version]. http://www.worldvaluessurvey.org

Johnson, E. C., Meade, A. W., & DuVernet, A. M. (2009). The role of referent indicators in tests of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 642–657. https://doi.org/10.1080/10705510903206014

Joshanloo, M. (2016). Religiosity moderates the relationship between negative affect and life satisfaction: A study in 29 European countries. *Journal of Research in Personality*, *61*, 11–14. https://doi.org/10.1016/j.jrp.2016.01.001

Jowell, R. (1998). How comparative is comparative research? *American Behavioral Scientist*, *42*(2), 168–177. https://doi.org/10.1177/0002764298042002004

Kaplan, D., Kim, J. S., & Kim, S. Y. (2009). Multilevel latent variable modeling: Current reaserch and recent developments. In R. Millsap & A. Maydeu-Olivares (Eds.), *The SAGE Handbook of Quantitiative Methods in Psychology*. SAGE. https://doi.org/10.4135/9780857020994

Lemos, C. M., Gore, R. J., Puga-Gonzalez, I., & Shults, F. L. (2019). Dimensionality and factorial invariance of religiosity among Christians and the religiously unaffiliated: A cross-cultural analysis based on the International Social Survey Programme. *PLOS ONE*, *14*(5). https://doi.org/10.1371/journal.pone.0216352

Liefbroer, A. C., & Rijken, A. J. (2019). The association between Christianity and marriage attitudes in Europe. Does religious context matter? *European Sociological Review*, *35*(3), 363–379. https://doi.org/10.1093/esr/jcz014

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, *23*(3), 524–545. https://doi.org/10.1037/met0000113

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

McAndrew, S., & Voas, D. (2011). Measuring religiosity using surveys. *Survey Question Bank: Topic overview*, *4*(2), 1–15.

McGuire, M. B. (2008). *Lived religion: Faith and practice in everyday life*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195172621.001.0001

Meuleman, B., & Billiet, J. (2011). Religious involvement: Its relation to values and social attitudes. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 173–206). Routledge.

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research*, *3*(1), 111–130. https://doi.org/10.21500/20112084.857

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge. https://doi.org/10.4324/9780203821961

Montgomery, K. A., & Winter, R. (2015). Explaining the religion gap in support for radical right parties in Europe. *Politics and Religion*, *8*(2), 379–403. https://doi.org/10.1017/s1755048315000292

Muthén, B., & Asparouhov, T. (2013). BSEM measurement invariance analysis. Mplus Web Notes: No17. https://www.statmodel.com/examples/webnotes/webnote17.pdf

Muthén, B., & Asparouhov, T. (2014). IRT studies of many groups: the alignment method. *Frontiers in Psychology*, *5*, 978. https://doi.org/10.3389/fpsyg.2014.00978

Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods & Research*, *47*(4), 637–664.

Muthén, L. K., & Muthén, B. (1998–2019). *Mplus User's guide*. Muthén & Muthén.

Norris, P., & Inglehart, R. (2004). *Sacred and Secular*. Cambridge University Press. https://doi.org/10.1017/cbo9780511791017

Pew Research Center. (2016). The gender gap in religion around the world. https://www.pewforum.org/2016/03/22/the-gender-gap-in-religion-around-the-world/

Pokropek, A., Davidov, E., & Schmidt, P. (2019). A monte carlo simulation study to assess the appropriateness of traditional and newer approaches to test for measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(5), 724–744. https://doi.org/10.1080/10705511.2018.1561293

Remizova, A., Rudnev, M., & Davidov, E. (2022). In search of a comparable measure of generalized individual religiosity in the World Values Survey. *Sociological Methods and Research*, 1–33.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software*, *48*(2), 1–36.

Saroglou, V., & Cohen, A. B. (2013). Cultural and cross-cultural psychology of religion. In R. F. Paloutzian & C. L. Park (Eds.), *Handbook of the psychology of religion and spirituality* (pp. 330–354). The Guilford Press.

Schwartz, S. H. (2012). Values and religion in adolescent development: Cross-national and comparative evidence. In G. Trommsdorff & X. Chen (Eds.), *Values, Religion, and Culture in Adolescent Development* (pp. 97–122). Cambridge University Press. https://doi.org/10.1017/cbo9781139013659

Siegers, P. (2011). A multiple group latent class analysis of religious orientations in Europe. In E. Davidov, P. Schmidt, & J. Billiet (Eds.), *Cross-cultural analysis: Methods and application* (pp. 387–413). Routledge.

Smith, T. W. (2017). Using the total survey error paradigm to improve cross-national research on religion. In R. Finke & C. D. Bader (Eds.), *Faithful measures: New methods in the measurement of religion* (pp. 167–201). New York University Press. https://doi.org/0.18574/nyu/9781479875214.001.0001

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*(1), 78–107. https://doi.org/10.1086/209528

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Pearson.

Tarakeshwar, N., Stanton, J., & Pargament, K. I. (2003). Religion: An overlooked dimension in cross-cultural psychology. *Journal of Cross-Cultural Psychology*, *34*(4), 377–394. https://doi.org/10.1177/0022022103034004001

Triandafyllidou, A., & Magazzini, T. (2021). *The Routledge handbook on the governance of religious diversity*. Routledge. https://doi.org/10.4324/9781003083405

van de Schoot, R., Kluytmans, A., Tummers, L., Lugtig, P., Hox, J., & Muthén, B. (2013). Facing off with Scylla and Charybdis: A comparison of scalar, partial, and the novel possibility of approximate measurement invariance. *Frontiers in Psychology*, *4*, 770. https://doi.org/10.3389/fpsyg.2013.00770

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70. https://doi.org/10.1177/109442810031002

Wilson, B. (1998). Secularization thesis: Criticism and rebuttals. In R. Laermans (Ed.), *Secularization and social integration: Papers in honor of Karel Dobbelaere, Leuven* (pp. 45–65). Leuven University Press.

Woodberry, R. D., Smith, C., & Scheitle, C. P. (2020). The Wiley Blackwell companion to sociology. In G. Ritzer & W. W. Murphy (Eds.), *Sociology of religion* (2nd, pp. 224–249). Wiley. https://doi.org/10.1002/9781119429333

Yuan, K.-H., & Chan, W. (2016). Measurement invariance via multigroup SEM: Issues and solutions with chi-square-difference tests. *Psychological Methods*, *21*(3), 405–426. https://doi.org/10.1037/met0000080