International Journal of Social Research Methodology

ISSN: 1364-5579 (Print) 1464-5300 (Online) Journal homepage: https://www.tandfonline.com/loi/tsrm20

No two ageisms are the same: testing measurement invariance in ageism experience across Europe

Romana Trusinová

To cite this article: Romana Trusinová (2014) No two ageisms are the same: testing measurement invariance in ageism experience across Europe, International Journal of Social Research Methodology, 17:6, 659-675, DOI: 10.1080/13645579.2013.823003

To link to this article: https://doi.org/10.1080/13645579.2013.823003

International Journal of Social Research Methodology, 2014
Vol. 17, No. 6, 659–675, http://dx.doi.org/10.1080/13645579.2013.823003

Routledge
Taylor & Francis Group

No two ageisms are the same: testing measurement invariance in ageism experience across Europe

Romana Trusinová*

Institute of Sociology of the Academy of Sciences of the Czech Republic, Jilská 1, Prague 1, 110 00, Czech Republic

(*Received 9 September 2012; final version received 3 July 2013*)

The article focuses on the topic of measurement invariance (equivalence) testing in comparative research. Measurement invariance is the basic requirement for comparative data and should be addressed by researchers before any international comparison begins. In the article, multiple-group confirmatory factor analysis is first performed to describe three methods of measurement invariance testing (configural, metric, and scalar). Subsequently, this procedure is applied to the case of measuring experience with ageism against seniors in the quantitative European Social Survey, Round 4. Measurement invariance in the concept is tested across 29 states participating in the survey. Hypotheses about the sources of the lack of measurement invariance between some countries are then tested by comparing the parameters of the models identified for different groups of European states. Finally, evidence of international comparability of items measuring experience with ageism against seniors is summarized and implications for other areas of comparative social research are discussed.

Keywords: measurement invariance; ageism; European Social Survey

Introduction

Measurement invariance is the basic requirement for data used in comparative research. Its absence may result in misinterpretation of results when differences caused by various measurement errors are interpreted as cultural differences in opinions, values, or experiences. The question of comparability of (not only) international data is answered by applying the method of measurement invariance testing (Comsa, 2010; Robert, Lee, & Chan, 2006; Schmitt & Kuljanin, 2008; Steenkamp & Baumgartner, 1998; Steinmetz, Schmidt, Tina-Booh, Wieczorek & Schwartz, 2009; Vandenberg & Lance, 2000). The primary goal of this paper is to present this method, emphasize its importance, and apply it on the issue of ageism in the quantitative European Social Survey, Round 4 (ESS4). The findings will be used (1) to decide which of the 29 countries taking part in the ESS4 can be included in a meaningful comparative analysis of ageism in Europe, and (2) to discuss the cultural sources of the lack of measurement invariance in age discrimination in the other countries. The results will also serve to evidence the quality of a highly-standardized international survey such as the ESS4 and to illustrate the pitfalls of comparative analysis in the study of other types of discrimination in society.

*Email: romana.trusinova@soc.cas.cz

The theory section of this article presents the concept of ageism and the method of measurement invariance testing. Subsequently, hypotheses about the form of measurement invariance in the specific study of ageism against seniors are formulated. The empirical section characterizes the data and models tested, the testing criteria, and the results. The concluding discussion both summarizes the main findings and describes its implications for further research in the field of discrimination based on (not only) age.

The concept of ageism

The concept of ageism was first used by American psychiatrist and gerontologist Butler (1969) in the late 60s of the last century. Over the following forty years, interest in the study of this phenomenon grew steeply and ageism as a word gradually became as well-known as sexism or racism. According to a simple definition by American sociologist Palmore, ageism is '*any prejudice or discrimination against or in favor of an age group*' (Palmore, 1990, p. 4). Although ageism means discrimination on the basis of any age (including young age), most studies in this area focus on the dimension of discrimination against seniors. So does the present article.

Nowadays, the topic of ageism is more relevant than ever before because of changes of demographic structure in developed countries. People live longer and the relative proportion of older people is growing. At the same time, people are healthier and more active due to advances in medicine and changes in lifestyle. In spite of that, the society does not seem to be abandoning some stereotypical notions of seniors. In this context, growing interest in ageism studies results in increased efforts to build empirical evidence about its extent, form, and causes. Attempts to explain the causes of ageism have had mixed results so far. Demographic, value, psychological, media, political, economic, and institutional factors are mentioned among the possible causes. Comparative research represents a suitable way of studying the causes of ageism because these causes often root in culture (Schoenberg & Lewis, 2005).

Although comparative international studies of ageism have been conducted before (e.g. Active Age, 2005; Eurobarometer, 2002), none of them focused on testing the assumption about measurement invariance in people's experiences with ageism across Europe which were being described. Given the highly symbolic nature of the concept of ageism, one can reasonably doubt that experiences with acts of prejudice or discrimination are both perceived and reported equivalently by respondents from different countries. What is considered grave disrespect and discrimination by one people can be perceived as harmless and irrelevant by members of another. It is precisely the issue of equal understanding and reflection of ageism experiences in the survey's response scales that can be addressed by the method of measurement invariance testing.

Measurement invariance

Measurement model

The argument that measurement invariance ought to be tested across groups (nations) builds on the assumption that the observed value of the concept depends

not only on its latent (actual) value but also on other parameters of the measurement model. The model characterizes the relationship between the latent concept which we intend to measure and the observed values of manifest indicators (Saris & Gallhofer, 2007, pp. 329–335; Steenkamp & Baumgartner, 1998; Wicherts, Dolan, & Hessen, 2005). Thus, before describing the procedure of invariance testing, I shall clarify the different parts of the measurement model. A standard measurement model takes the form of Formula 1:

$$y = \tau + \lambda \mathrm{CP} + \xi \tag{1}$$

where y = observed value, CP = actual value of the concept we want to measure, τ = intercept term, or the y value for CP = 0, λ = regression slope, or regression coefficient, and ξ = random error caused by other factors.

The model of measuring the concept by means of a questionnaire question can be depicted as a regression line (Figure 1). This is a function of response to a question that represents the given latent concept ideally. Intercept $\tau = 0$, i.e. when the concept equals zero, respondents choose zero on the response scale indicating the concept. At the same time, regression coefficient $\lambda = 1$, i.e. respondents increase their answers on the response scale by one every time the value of the concept grows by one.

However, the measurement model may take very different forms as well. Various combinations of the τ and λ values may result in a situation depicted in Figure 2. Bold lines represent the regression lines of the concept's measurement by the indicator in four theoretical groups of respondents (G1–G4). Dotted lines show how the same actual value of the concept may result in four highly divergent observed values y in four different measurement models.

If we find a situation in the data where the regression line of the relationship between the concept and the indicator differs across groups, then we can assume that the given question does not measure the same concept in all the groups, or at least it does not measure it in the same ways. Thus, it would be misleading to merely compare the mean values of answers to the question across the group.

Measurement invariance testing

Measurement invariance occurs if there are no significant differences between the measurement models in each group of respondents (here, people living in different

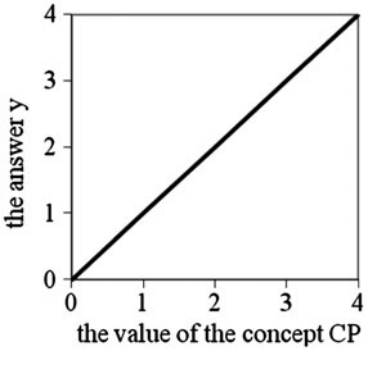


Figure 1. The regression line for $\tau = 0$ and $\lambda = 1$.

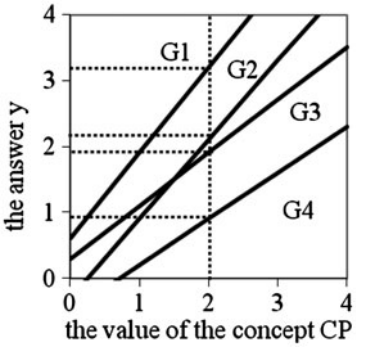


Figure 2. The regression line for different non-zero values of τ and λ.

countries), i.e. observed values y represent the values of concept CP similarly in all groups (Robert et al., 2006; Saris & Gallhofer, 2007, pp. 329–335; Schmitt & Kuljanin, 2008; Steenkamp & Baumgartner, 1998; Steinmetz et al., 2009; Vandenberg & Lance, 2000). At least three indicators measuring the concept are necessary for measurement invariance testing. These three indicators of a single concept allow us to formulate a system of equations (Formulas 2–4) which can be solved through confirmatory factor analysis.

$$y_1 = \tau_1 + \lambda_{11}\mathrm{CP} + \xi_1 \tag{2}$$

$$y_2 = \tau_2 + \lambda_{21}\mathrm{CP} + \xi_2 \tag{3}$$

$$y_3 = \tau_3 + \lambda_{31}\mathrm{CP} + \xi_3 \tag{4}$$

Thus, to test measurement invariance means to apply this system of equations and answer three important questions: (1) Are we measuring the same concept in all groups? (2) Are the measurement parameters equal across nations?, and (3) Can differences in manifest scores be interpreted as actual differences in the latent concept? (Steinmetz et al., 2009).

Measurement invariance in ageism against seniors: the model and research hypotheses

The ESS4, measured the concept of ageism experience by means of three questions. All three were retrospective and inquired about the frequency of respondents' past-year experiences with ageism. A response scale from 0 to 4 was used where 0 meant 'never' and 4 'very often.' The questions were worded as follows:

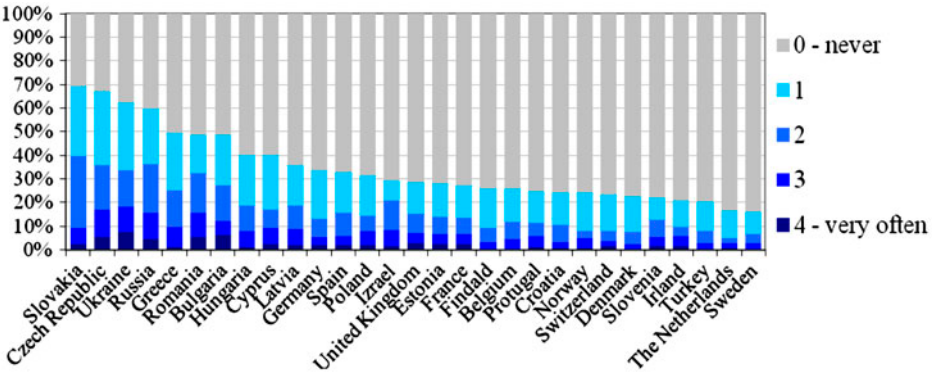
> Q1: Using this card please tell me how often, in the past year, anyone has shown prejudice against you or treated you unfairly because of your age?
>
> Q2: And how often, if at all, in the past year have you felt that someone showed you a lack of respect because of your age, for instance by ignoring or patronizing you?
>
> Q3: In particular, how often in the past year has someone treated you badly because of your age, for example by insulting you, abusing you, or refusing you services?

The three questions on ageism experience were asked of all respondents, irrespective of their age. Indeed, according to the definition of ageism, this type of discrimination may affect a person of any age. In this text, however, attention is paid specifically to ageism against old age. This is because ageism against young age (or other age groups) can be expected to have different causes and manifestations and, overall, to be conceptually different from ageism against seniors. If the research sample was not limited to respondents of one age group, the study would be too complicated and its findings difficult to interpret. Moreover, as demonstrated by prior studies, measurement invariance may be problematic not only between cultures but also between age groups or other social categories, even for questions that are much less associated with age (Anderson, Lievens, Dam, & Born, 2006; Steinmetz et al., 2009; Tucker, Ozer, Lyubomirsky, & Boehm, 2006; Wicherts et al., 2005).

If we want to study ageism against old age, we should only include in the analysis answers by respondents we consider 'old.' The quotation marks are to suggest the symbolic nature of the notion of old age – its strong dependence on both cultural context and subjective perception at the individual level. That said, I consider limiting the research sample by age to be the purest possible solution. Seventy years will be chosen as the cutoff age for including respondents in the sample, thus for their being perceived as seniors. This value was chosen based on recommendations arising from sociologists' discussion during the preparations of the ESS from which the present article takes its data; the ESS used the cutoff value of 70 years in questions mapping respondents' attitudes to seniors. The fact that this value is not selected randomly or by error is confirmed by analyses of other questions asked in the ESS. In the different countries, the average age at which respondents thought old age begins was around 63 years. In no European country was old age perceived to begin at a higher age than 70 (the highest age, 68 years, was stated by Greeks). Thus, a 70-year-old person can be expected to be considered old by most of Europe's population and therefore, that person's experience with discrimination on the basis of age can be understood as discrimination against old age. This justifies using a unified cutoff value of 70 years for including respondents in a study of measurement invariance in ageism against seniors.

Let us now return to the three above-mentioned indicators of ageism against seniors used in this study. A frequency chart (Graph 1) was plotted for the second question on experiences with ageism. The data presented in the chart are already limited to respondents aged 70 or older. It is clear that seniors across Europe reflect highly divergent experiences with acts of a lack of respect (hereinafter referred to as disrespect), from almost 70% disrespected seniors in Slovakia and the Czech Republic to 17% in The Netherlands and 16% in Sweden. These figures alone beg some interpretation. However, as noted by Saris and Gallhofer (2007, p. 339), among others, international comparison derived from answers to a single item without any information about measurement invariance is at least courageous. Therefore, this article uses a different method of international comparison based on (1) measurement invariance testing and (2) comparison of latent means among states with demonstrated measurement invariance.

Three manifest variables measured by the above questions will serve as reflective indicators of the latent concept of perceived ageism against seniors in society. Question 1 will be shortly referred to as level of experience with prejudice, Question 2 as level of experience with disrespect, and Question 3 as level of



Graph 1. Experiences with acts of disrespect based on age, international comparison.

experience with age discrimination. The relationship between these indicators and the concept is modeled in Figure 3.

Hypothesis 1, as tested in this article, is as follows: *Ageism measurement is not invariant in all 29 states.* The related question is which of the participating countries can be considered comparable. Empirical texts on measurement invariance testing often conclude by confirming or rejecting this hypothesis without attempting to explain any substantive sources of the lack of measurement invariance (e.g. Byrne, 1994; Comsa, 2010). On one hand, such procedure is understandable given the possible complex sources of divergence between the parameters of the measurement model in different cultures. On the other hand, some readers may perceive an article such a conclusion as unfinished story because to explain the lack of measurement invariance might be the most important part of the story. Therefore, in this article, I will attempt to map some reasons of the lack of measurement invariance in ageism across countries and draw some implications not only for ageism studies but also for studies of other forms of discrimination.

In an inspiring article, Robert et al. (2006) describe 11 factors that may affect measurement invariance, distinguishing between four groups: culture, language and translation, respondent's membership in organizations, and interview context. This

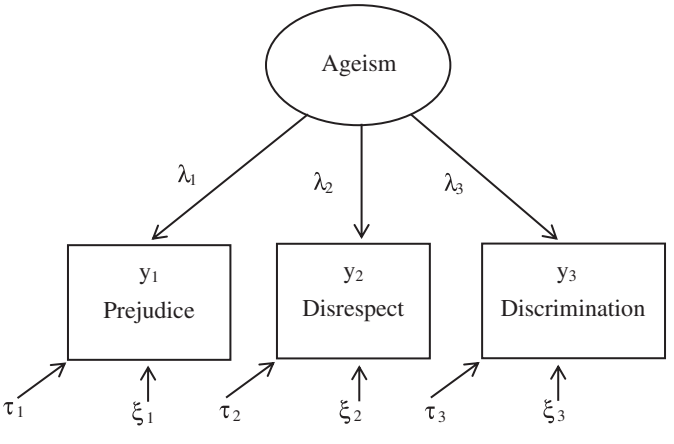


Figure 3. A model of measuring ageism in society.

article only pays attention to cultural factors influencing measurement invariance because the empirical section focuses on testing international measurement invariance (see Robert et al., 2006, for details about other factors influencing measurement invariance).

Different meanings of the measured concept in different cultures can be expected to represent an important cultural factor of the lack of measurement invariance. What people in one country consider patronizing, disrespectful, or unfair may seem normal or irrelevant to another country's citizens. This gives rise to the second hypothesis of this article, namely that *the lack of measurement invariance in ageism experience results from higher sensitivity of people living in countries with longer democratic traditions in identifying prejudice or disrespect in people's actions.* Thus, irrespective of the total scores of the concept of ageism, I assume higher regression coefficients λ_1 and λ_2 among seniors from Western and Northern European countries, compared to those from Eastern (or even Central) Europe.

The last, third hypothesis also assumes cultural effects on measurement invariance. It assumes divergent effects of social desirability, which exists in all cultures but may take specific cultural forms (Robert et al., 2006). The hypothesis is worded as follows: *the lack of measurement invariance is also caused by the fact that in countries where the topic of demographic panic and intergenerational conflict lies in the centre of public attention, it is more socially desirable for seniors to report experiences with prejudice and disrespect.* Thus, we expect different results for countries where the topic of demographic aging and the often related issue of 'generational war' is more popular and where, as a result, complaints about disrespect for seniors belong to normal conversation repertoire. Therefore, I expect relatively higher values of intercepts τ_1 and τ_2 in these countries, compared to countries where this topic is less popular.

The question is which countries are more involved in the issue of intergenerational conflict. To simplify the answer, I will use the indicator of demographic forecast in individual countries. The issue of demographic aging can be expected to be more burning for countries which are predicted to experience the largest growth of seniors' proportion in the population structure in the following years (till 2050). These include countries in Central and Eastern Europe and the Iberian Peninsula (Eurostat, 2011).[1]

Data and methods

Data

I tested the above three hypotheses on data from the ESS4. The ESS,[2] a long-term project of social research, mainly consists of international quantitative surveys of values, attitudes, behavior, and sociodemographic and sociostructural characteristics of populations in European countries. The surveys are implemented every 2 years. The ESS4 took place from 2008 to 2009 in 29 states: Belgium, Bulgaria, Croatia, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Lithuania, The Netherlands, Norway, Poland, Portugal, Rumania, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, Ukraine, and the UK. Data were collected by means of face-to-face interviewing, mostly with paper questionnaires. It is the ESS4 which contained a special module of questions on 'Experiences and Expressions of Ageism,' including the above-mentioned three questions about experiences with different acts of ageism.

As stated in the previous section, the sample was limited to respondents aged 70 or older for the sake of conceptual clarity. By limiting the sample, I obtained a more compact data-set of a total of 6857 seniors, i.e. an average of 235 seniors per country. The number of senior respondents differs between countries, from 120 in Croatia to 475 in Portugal, depending on the total number of respondents and the demographic structure of the sample. Data-sets of this size are perfectly sufficient for testing the required structural model of ageism.

Characterizing the models

The method of multi-group confirmatory factor analysis was used to test measurement invariance. The model depicted in Figure 3 was tested on all 29 countries taking part in the ESS4. Three types of measurement invariance characterized by increasingly restrictive models were tested (Milfont & Fischer, 2010; Schmitt & Kuljanin, 2008; Steenkamp & Baumgartner, 1998; Steinmetz et al., 2009; Vandenberg & Lance, 2000):

- Configural invariance – the basic type of invariance which occurs if the same model fits the data in all participating countries. Here, configural invariance means that the three manifest variables can be considered indicators of a single latent concept of ageism in all the countries. This model's validity is a necessary, but not sufficient condition for comparability of the concept across states.
- Metric invariance – occurs if the condition of configural invariance is satisfied and, at the same time, regression coefficients λ_1, λ_2, and λ_3 are equal across countries. While this type of invariance does not legitimize comparing means or scores for individual manifest variables, it is a necessary condition for comparing relationships across countries as well as for comparing latent means (Saris & Gallhofer, 2007, pp. 342–346; Steinmetz et al., 2009). By latent means we understand mean values of the latent concept calculated from the manifest means of indicators and known values of slopes and intercepts. The latent mean is a relative quantity because it is constrained to 0 for one group (the reference group). Latent means thus help us compare the value of the concept between the groups (countries) under investigation even in the absence of the third type of invariance measurement, scalar invariance (Byrne, 2001).
- Scalar invariance – occurs if the condition of metric invariance is satisfied and, at the same time, intercepts τ_1, τ_2, and τ_3 are equal across states. It is only in the presence of scalar invariance that the same measurement models apply to the different countries and, in turn, different values of both the observed concept and the different manifest variables can be interpreted as substantial differences in the value of the latent concept of ageism (making means comparable).

The relatively strict requirement of supporting all three types of measurement invariance often results in rejecting the hypothesis of measurement invariance of across countries (Milfont & Fischer, 2010; Robert et al., 2006; Schmitt & Kuljanin, 2008; Steinmetz et al., 2009; Vandenberg & Lance, 2000). For this purpose, researchers have attempted to prove that partial invariance is a sufficient condition

of data comparability (Baumgartner & Steenkamp, 1998; Byrne, Shavelson, & Muthén, 1989). Partial invariance occurs if most of the model's parameters are invariant across the groups (states) analyzed, while some regression coefficients and intercepts can be omitted. Based on these authors' arguments, I will consider the presence of at least two invariantly measured indicators of the concept of ageism as a sufficient condition of comparability.

I will test the second and third hypotheses about the sources of any lack of measurement invariance by attempting to identify separate models of invariance measurement for different groups of European states. If such separate models are identified, their parameters will be compared. Different values of regression coefficients or intercepts in ageism measurement models across Europe will be used to interpret the possible sources of the lack of measurement invariance and, in turn, to confirm or reject the second and third hypotheses.

Criteria for testing and interpreting the models

The LISREL software was used for data analysis. First, covariance matrixes, means, and standard errors of the variables were computed for each country. Based on these input data, the models to test were specified.

Since the latent variable in the model does not have its own scale, such a scale must be set by constraining one parameter λ to 1 and one parameter τ to 0. Then, the latent concept's scale is equal to that of the given indicator. The decision which indicator is the most suitable for constraining may affect the results of partial invariance testing across states, for example when exactly one non-equivalent indicator is constrained (Schmitt & Kuljanin, 2008; Steinmetz et al., 2009; Vandenberg & Lance, 2000).

In the following analysis, I will constrain each latent concept scale to equal variable y_3, i.e. the question about experience with age discrimination. I chose so because the wording of Question 3 includes the most specific examples which can be expected to be met with the most consistent understanding by respondents across nations. While this thesis is intuitive, it can be assumed that respondents from different nations are more likely to agree on what is or is not 'insult, abuse or refusal of services' (see Question 3), compared to what is or is not 'showing prejudice or unfair treatment' (see Question 1) or 'ignoring or patronising' (see Question 2). At the same time, Question 3 inquires about a very clear and extreme act of ageism in society. While Questions 1 and 2 rather measure the emotional perceptions of hostility, the third one identifies experiences with physically apparent discriminatory actions. Thus, any higher levels of parameters λ_1, λ_2 in some countries will be interpreted as higher sensitivity in identifying actions crossing the line set by Questions 1 and 2.

To confirm or reject tested models is a complex decision and different authors do not agree on the most suitable procedure (some of these discussions are summarized, for example, by Arbuckle & Wothke, 1999 or Byrne, 2001). There is a large number of criteria for testing models in confirmatory factor analysis and a large number of recommendations on cutoff values for the different parameters (Arbuckle, 1999; Bollen, 1989; Byrne, 1994; Milfont & Fischer, 2010; Schmitt & Kuljanin, 2008; Vandenberg & Lance, 2000). It is beyond the possibilities and ambitions of this article to contribute to this methodological discussion. Therefore, in the following, I merely summarize the rules I derived from other authors' recommendations and used for the tests presented in the empirical section.

To decide whether the tested models fit well, I will use several criteria in line with consensual recommendations for the method of confirmatory factor analysis. The basic requirement for a model to be accepted lies in low level of the χ^2 statistic and statistical insignificance of this test. The χ^2 test works with the zero hypothesis that the covariance matrix of observed values is equal to the matrix estimated on the basis of the model. If the test is statistically significant, the hypothesis is rejected and an alternative hypothesis about lack of fit is accepted (residue values are too high). It is even more suitable to consider the level of the χ^2 statistic relatively to the number of degrees of freedom, while this text will work with a cutoff value of $\chi^2/df = 3$ (Arbuckle, 1999).

It has been debated whether the χ^2 statistic represents a sufficient criterion for confirming or rejecting the model, especially given its dependence on sample size N (Bollen, 1989). Therefore, I will use other goodness-of-fit indices as well. The root mean square error of approximation (RMSEA), currently one of the most frequently used goodness-of-fit indices, will serve as another criterion of a model's overall validity. Its value should not exceed the recommended limit of .05, or .08 at maximum (Arbuckle, 1999).

I will also check two indicators comparing the model with a baseline model, i.e. one assuming no correlation between manifest variables. The Tucker Lewis index or non-normed fit index (NNFI) and the comparative fit index (CFI) are examples of such indicators. They range between zero and one, and models with values close to one are considered as fitting the data. In this article, I will work with the cutoff value of .95 (Arbuckle, 1999).

Furthermore, I will identify the levels of modification indices. Modification indices help us identify any improvement in the tested model achieved after including a specific relationship in the model by estimating decreases in the level of the χ^2 statistic (Arbuckle, 1999). In the following text, models with no modification indices of six or more will be considered good.[3]

Results: ageism measurement invariance in the ESS4

Testing the hypotheses about lack of measurement invariance across Europe

In the first step towards answering the research question, I tested the configural invariance model by freeing all model parameters (except those constraining the scale of the latent variable). In the second step, I constrained the regression coefficients λ as invariant across states (metric invariance), and in the third step, I also constrained the intercepts τ (scalar invariance). As the results in Table 1 demonstrate, only the configural invariance model fits the data across all countries. In contrast, for metric and scalar invariance models yield unacceptable results.

Based on this finding, the first hypothesis can be accepted, namely that ageism measurement is not invariant across all participating states. In order to determine

Table 1. Results of the invariance models for all 29 states.

	χ^2	df	χ^2/df	P	RMSEA
Configural	34.5	45	.76	.87	.000
Metric	318	54	5.9	.000	.145
Scalar	650	101	6.4	.000	.146

Table 2. The resulting partial metric and scalar invariance models for selected countries.

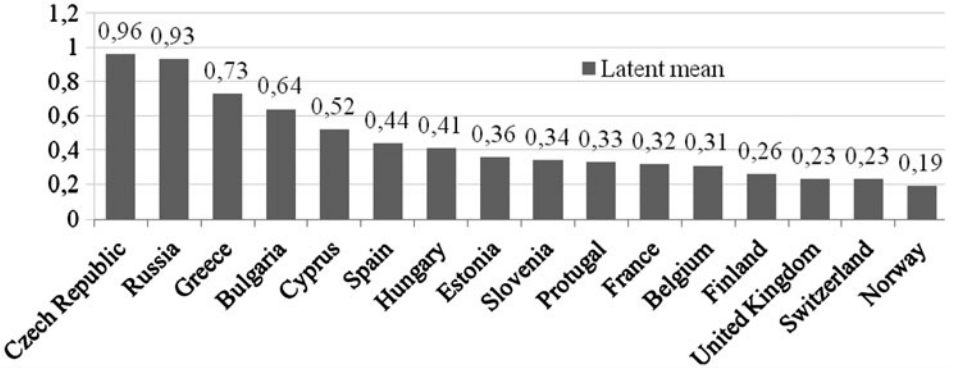
	χ^2	d*f*	χ^2/df	*P*	RMSEA
Partial metric invariance, 16 European states	27.9	21	1.33	.14	.037
Partial scalar invariance, 12 European states	39.9	31	1.29	.13	.034

which countries can be considered comparable, the models were adjusted along two lines: (1) by freeing some model parameters, evidence of at least partial measurement invariance will be obtained; and (2) some states were excluded from the analyses because their measurement models are fundamentally different. By applying these steps gradually,[4] I eventually identified a partial metric invariance model[5] and a partial scalar invariance model[6] for 16 and 12 countries, respectively. The results of these models are presented in Table 2.

According to the results, which countries can be used to compare relationships between variables and to compare latent means? In other words, which countries fulfill the conditions of partial metric invariance? They are as follows: Belgium, Bulgaria, Cyprus, the Czech Republic, Estonia, Finland, France, Greece, Hungary, Norway, Portugal, Russia, Slovenia, Spain, Switzerland, and the UK.

And which countries can be compared on the basis of mean values of ageism experience or scores for each question (i.e. which fulfill the conditions of partial scalar invariance)? They include all of the countries mentioned in the preceding paragraph except Greece, Norway, Russia, and the UK.

Using the latent means of the concept of ageism towards seniors in society which can be computed in LISREL, we can finally obtain an empirically more stable alternative to the original Graph 1. Compared to Graph 1, it includes less states and the concept of latent means it works with is more difficult to interpret. However, it can be verified methodologically, which makes its interpretation more persuasive. Graph 2 informs us that the level of ageism toward seniors is the highest in the Czech Republic and Russia; followed by a considerable distance by Greece and Bulgaria. At the other extreme, there are Norway, Switzerland, and the UK. Generally speaking, the problem of ageism toward seniors is the most prevalent in countries with a communist past. In the next step, data from the 12 countries fulfill-



Graph 2. Comparing latent means of ageism in countries fulfilling the requirements of partial metric invariance.

ing the conditions of partial scalar invariance can be used for a more precise analysis of the cultural causes of ageism.

Testing the hypotheses about the cultural sources of the lack of measurement invariance

The states whose ageism measurement is not equivalent to measurements in the 12 European countries included in the partial scalar invariance model in Table 2 are as follows: Croatia, Denmark, Germany, Greece, Ireland, Israel, Lithuania, The Netherlands, Norway, Poland, Rumania, Russia, Slovakia, Sweden, Turkey, Ukraine, and the UK. Since these countries have different ageism measurement models, it makes sense to ask about the causes of such lack of measurement invariance.

We can get some answers by estimating the parameters of separate measurement invariance models for the states mentioned in the preceding paragraph and comparing them with the parameters of the partial scalar invariance model in Table 2. Two further models of partial scalar invariance between those countries can be identified (Table 3). The first one includes Denmark, Germany, Norway, Sweden, and the UK. The other one confirms measurement invariance for Croatia, Greece, Lithuania, Poland, Rumania, and Russia. The remaining countries (Ireland, Israel, The Netherlands, Slovakia, Turkey, and Ukraine) have such highly specific parameters of ageism measurement models that they correspond with none of the three models identified.

In order to confirm or reject the hypotheses about the effects of cultural background on sensitivity towards acts of prejudice and disrespect (Hypothesis 2) and about social desirability affecting the intercept (Hypothesis 3), it is necessary to identify the specific values of parameters in the different models. They are stated in regression equations (5)–(13). The equations are organized in three blocks which correspond to the three groups of countries whose models of ageism measurement models are being compared. Each block contains equations representing the relationship of the three manifest variables with the latent variable of ageism experience:

Europe 12 (Belgium, Bulgaria, Cyprus, the Czech Republic, Estonia, Finland, France, Hungary, Portugal, Slovenia, Spain, and Switzerland)

$$y_{\text{prejudice}} = 0.02 + 1.28 \times \text{ageism} + \xi_1 \tag{5}$$

$$y_{\text{disrespect}} = -0.03 + 1.49 \times \text{ageism} + \xi_2 \tag{6}$$

$$y_{\text{discrimination}} = \text{ageism} + \xi_3 \tag{7}$$

Table 3. The resulting partial scalar invariance models for selected countries.

	χ^2	df	χ^2/df	P	RMSEA
Partial scalar invariance: Denmark, Germany, Norway, Sweden, and the UK	19.6	14	1.4	.15	.039
Partial scalar invariance: Croatia, Greece, Lithuania, Poland, Rumania, and Russia	8.1	14	.59	.88	.000

Northern Europe + Germany (Denmark, Germany, Norway, Sweden, and the UK)

$$y_{\text{prejudice}} = 0.02 + 1.66 \times \text{ageism} + \xi_1 \tag{8}$$

$$y_{\text{disrespect}} = 0.01 + 2.03 \times \text{ageism} + \xi_2 \tag{9}$$

$$y_{\text{discrimination}} = \text{ageism} + \xi_3 \tag{10}$$

Eastern and South-Eastern Europe (Croatia, Greece, Lithuania, Poland, Rumania, and Russia)

$$y_{\text{prejudice}} = 0 + 0.88 \times \text{ageism} + \xi_1 \tag{11}$$

$$y_{\text{disrespect}} = -0.07 + 1.31 \times \text{ageism} + \xi_2 \tag{12}$$

$$y_{\text{discrimination}} = \text{ageism} + \xi_3 \tag{13}$$

The regression equations are expressed graphically in Figures 4 and 5. Figure 4 depicts the linear regression function of answers to the question on level of experience with prejudice. Figure 5 depicts the same function for the question on level of experience with disrespect. The *x*-axis shows the latent concept of ageism whose scale is constrained to equal the variable measuring experience with age discrimination. The *y*-axis shows estimated values of indicators in line with the models identified.

The above figures and equations make it clear that the values of regression coefficients λ_1 and λ_2 are higher for Nordic states and Germany, compared to the other countries. In contrast, these parameters are statistically significantly lower in Eastern and South-Eastern European countries, compared to the rest of Europe. These results confirm the second hypothesis. Given the same overall level of ageism in society, Nordic peoples reflect experience with acts of prejudice and disrespect more strongly, while Eastern and South-Eastern European nations are rather tolerant towards these acts. However, the second hypothesis is accepted as non-universal.

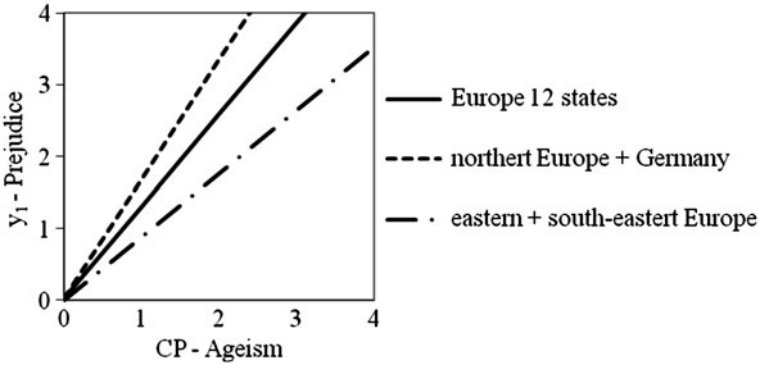


Figure 4. Comparing the regression lines for Question 1, experience with prejudice.

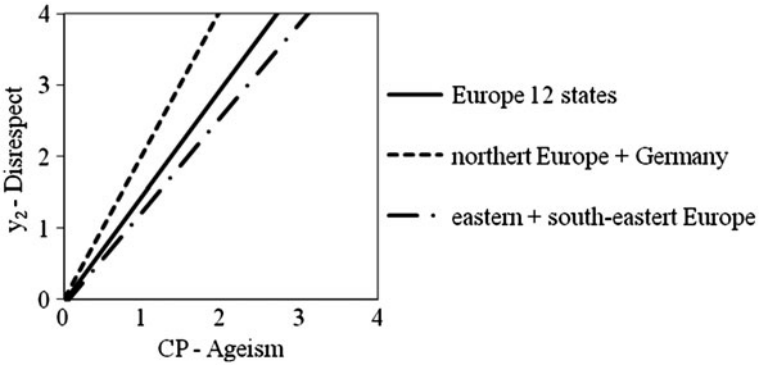


Figure 5. Comparing the regression lines for Question 1, experience with disrespect.

Some nations of Western and Northern Europe are more sensitive to acts of ageism, but this regularity does not apply elsewhere, and answers by people from different countries can be compared irrespective of geographic region.

The third hypothesis assumed different values of intercepts across Europe. While the values of τ_1 and τ_2 are statistically significantly different between the three above-mentioned models, Figures 4 and 5 illustrate that the differences are not substantively large and thus interesting for interpretation. Thus, I have rejected the third hypothesis about the effects of social desirability on the intercepts.

Discussion

Discussion of the results

The goal of this paper was to test measurement invariance in ageism against seniors across 29 states in the ESS4 survey. The test results showed that the three questions about seniors' experiences with ageism can be considered as indicators of a single concept of ageism in all 29 states participating in the ESS4. However, the tests of metric and scalar invariance revealed substantial differences in the parameters of measurement models between the states observed. Thus, the basic finding of this research study is that answers to the three questions cannot be simply compared across all the countries because their measurement is not fully internationally invariant.

Partial metric invariance was identified across 16 of the 29 states under observation. In these countries, the latent means of the concept of ageism are comparable, just like the relationships between its indicators and other variables. However, if we wanted to analyze the levels of ageism using a more in-depth comparative design, we should only use data from the above-defined group of 12 states where ageism measurement was found to support partial scalar invariance. For data from these countries, different scores of answers to the questions on experience with acts of ageism can be interpreted as different levels of ageism in these countries, rather than effects of unequal parameters of the model for measuring the concept.

In order to identify the cultural sources of the lack of measurement invariance through mutual comparison of model parameters, two further measurement models were created, one for Northern Europe and Germany and another one for a part of Eastern and South-Eastern Europe. The three models were compared to illustrate the fact that invariance can be limited by differences in the ways questions are

understood. Compared to other European nations, respondents from Nordic countries and Germany, all of which are countries with longer democratic traditions and high standards of living, seem to be more sensitive to acts of prejudice and disrespect because they relay more frequent experiences with symbolic acts of ageism for the same levels of ageism experience as such. In contrast, seniors from South-Eastern and Eastern Europe were found to be much less sensitive to these experiences. Differences between the models' regression coefficients can be logically interpreted as evidence of effects of cultural differences on the lack of ageism measurement invariance across Europe.

This research study has two types of limitations. The first one lies in its inability to satisfactorily and fully interpret the differences between ageism measurement models across all European states. The second disadvantage of this way of measurement invariance testing lies in the necessity to exclude non-invariant states from further analysis; a more favorable solution would be to adjust the data in line with information about specific differences between the measurement models. However, no methodology for insuring comparability by such ex-post data adjustment is available at the moment.

Implications for further research

The results of the analyses demonstrated that the problem of lack of measurement invariance may actually occur even in highly standardized international surveys like the ESS4. Since this is caused by the countries' different cultural backgrounds, one cannot rely on achieving a quick solution by improving the survey methodology in the preparatory stage of data collection. At present situation, researchers who realize the pitfalls of the lack of measurement invariance in comparative research can only work on testing measurement invariance in the questions before any international comparison begins.

A good news for further comparative research of ageism against seniors in Europe is that 12 countries have been defined where measurement of a concept as complex as ageism supports partial scalar invariance. The fact that this group includes geographically and culturally distant regions means that the study of the cultural context of ageism promises interesting interpretations.

The results of the above-presented analyses may also inspire researchers focusing on other types of discrimination in intercultural comparison. Sensitivity for discrimination on the basis of gender, ethnicity, etc. can be expected to vary analogically to the way nations differ in sensitivity for age discrimination. Generally speaking, international comparison of opinions about other people's decency, fairness, and good conduct will always strongly depend on the cultural context and, i.e. the standard in which the respondent acts. In view of the above-presented results, nations with longer democratic traditions and higher standards of living can be expected to be more sensitive to expressions of discrimination between people. However, this logic cannot be apriori taken as a given because it may apply differently in different countries.

Acknowledgements

This article was supported by the research project "Creation of the Czech intersection of pan-european project of a research infrastructure European Social Survey", no. LM2010012, Ministry of Education, Youth and Sport of the Czech Republic.

Notes

1. As for other factors influencing measurement invariance, Robert et al. (2006) note that cultural origin may affect respondents' very strategies of applying the response scale. Many authors deal with this issue in the context of response styles. It has been demonstrated that some nations (e.g. Americans) tend to use the extremes of a scale while others (typically, Asian nations) are more inclined towards using the scale's center (see, e.g. Dolnicar & Grun, 2007, for a complex discussion of this topic in comparative research). However, in order to test this phenomenon, we need several questions measuring the same phenomenon by means of different scales. Such data on ageism are, unfortunately, not available as the three questions posed in ESS4 were accompanied by the same response scales.
2. See www.europeansocialsurvey.org for more information.
3. In contrast to the rest of the above-mentioned statistics, the NNFI, the CFI, and the modification index are not presented in the tables in the results section, but merely for the lack of space. For all models presented as valid, the values of these indices are in line with the criteria described above.
4. Among the 29 participating countries, several different metric or scalar invariance models can be identified for different combinations of countries. In each case, it is decisive which specific countries are excluded in the different steps and which parameters are freed. The resulting models presented are chosen to include the maximum number of countries.
5. The partial metric invariance model took the following form: the same model with one latent and three manifest variables was tested across the countries; in each case, λ_3 was constrained to 1 and τ_3 to 0 and, at the same time, λ_2 and τ_2 were set as invariant; λ_1 was also invariant for Bulgaria, Estonia, Finland, Cyprus, Spain, and Switzerland, τ_1 and τ_2 were free.
6. The model of partial scalar invariance took the following form: the same model with one latent and three manifest variables was tested across the countries; in each case, λ_3 was constrained to 1 and τ_3 to 0 and, at the same time, λ_2 and τ_2 were set as invariant; λ_1 was also invariant for Spain, Finland, and Cyprus, τ_1 was also invariant for Portugal and Slovenia, and all parameters (λ_1 and τ_1) were invariant for Estonia and Switzerland.

Notes on contributor

Romana Trusinová is a doctoral student of sociology at the Faculty of Social Sciences, Charles University in Prague. She works at the Department of Value Orientations in Society, Institute of Sociology of the Academy of Sciences of the Czech Republic. In her research, she specializes on intergenerational relations and ageism.

References

Active Age. (2005). *EU research on social sciences and humanities: Overcoming the barriers and seizing the opportunities for active ageing policies in Europe*. Retrieved from http://cordis.europa.eu/documents/documentlibrary/100124401EN6.pdf

Anderson, N., Lievens, F., Dam, K. V., & Born, M. (2006). A construct-driven investigation of gender differences in a leadership-role assessment center. *Journal of Applied Psychology, 91*, 555–566.

Arbuckle, J. L., & Wothke, W. (1999). *Amos 4.0 user's guide*. Chicago, IL: SPSS and Small Waters Corporation.

Baumgartner, H., & Steenkamp, J. B. E. M. (1998). Multi-group latent variable models for varying numbers of items and factors with cross-national and longitudinal applications. *Marketing Letters, 9*, 21–35.

Bollen, K. A. (1999). *Structural equations with latent variables*. New York, NY: Wiley.

Butler, R. (1969). Age-ism another form of bigotry. *Gerontologist, 9*, 243–246.

Byrne, B. (1994). Testing for factorial validity, replication, and invariance of a measuring instrument: A paradigmatic application based on the maslasch burnout inventory. *Multivariate Behavioral Research, 29*, 289–311.

Byrne, B. (2001). *Structural equation modelling with Amos: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum.

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*, 456–466.

Comsa, M. (2010). How to compare means of latent variables across countries and waves: Testing for invariance measurement. An application using eastern European societies. *Sociológia, 42*, 639–669.

Dolnicar, S., & Grunn, B. (2007). Assessing analytical robustness in cross-cultural comparisons. *International Journal of Culture, Tourism and Hospitality Research, 1*, 140–160.

Eurobarometer. (2002). *Discrimination in Europe*. Retrieved from http://ec.europa.eu/public_opinion/archives/ebs/ebs_168_exec.sum_en.pdf

Eurostat. (2011). *Population projections.* Retrieved from http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Population_projections

Milfont, T. L., & Fischer, R. (2010). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of Psychological Research, 3*, 111–121.

Palmore, E. B. (1990). *Ageism: Negative and positive*. New York, NY: Springer.

Robert, C., Lee, W. C., & Chan, K. (2006). An empirical analysis of measurement equivalence with the INDCOL measure of individualism and collectivism: Implications for valid cross-cultural inherence. *Personnel Psychology, 59*, 65–99.

Saris, W. E., & Gallhofer, I. N. (2007). *Design, evaluation, and analysis of questionnaires for survey research*. New York, NY: Wiley.

Schmitt, N., & Kuljanin, G. (2008). Measurement invariance: Review of practice and implications. *Human Resource Management Review, 18*, 210–222.

Schoenberg, N. E., & Lewis, D. C. (1988). Cross-cultural ageism. In L. B. Palmore, L. G. Branch, & D. K. Harris (Eds.), *Encyclopedia of ageism* (pp. 87–90). Binghamton: The Haworth Press.

Steenkamp, J. B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *The Journal of Consumer Research, 25*, 78–90.

Steinmetz, H., Schmidt, P., Tina-Booh, A., Wieczorek, S., & Schwartz, S. H. (2009). Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement. *Quality & Quantity, 43*, 599–616.

Tucker, K. L., Ozer, D. J., Lyubomirsky, S., & Boehm, J. K. (2006). Testing for measurement invariance in the satisfaction with life scale: A comparison of Russians and North Americans. *Social Indicators Research, 78*, 341–360.

Vandenberg, R. J., & Lance, Ch. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*, 4–70.

Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005). Stereotype threat and group differences in test performance: A question of measurement invariance. *Journal of Personality and Social Psychology, 89*, 696–716.