Survey Research Methods (2020)
Vol. 14, No. 4, pp. 351-364
doi:10.18148/srm/2020.v14i4.7369

ISSN 1864-3361
http://www.surveymethods.org

The comparability of measures in the ageism module of the fourth round of the European Social Survey, 2008–2009

Daniel Seddig
Faculty of Management, Economics and Social Sciences
University of Cologne, Germany

Dina Maskileyson
Faculty of Management, Economics and Social Sciences
University of Cologne, Germany

Eldad Davidov
Faculty of Management, Economics and Social Sciences
University of Cologne, Germany and
URPP "Social Networks" and Department of Sociology
University of Zurich, Switzerland and
The Minerva Center on Intersectionality in Aging
University of Haifa, Israel

Age is an important dimension that is used by people to categorize others. Age-based discrimination is directed toward specific age groups (young and old). In spite of their importance, attitudes toward the elderly have hardly been studied from a cross-country perspective. The fourth round of the European Social Survey (ESS) from 2008-2009 offers, for the first time, the opportunity to study ageism and attitudes toward age groups from a cross-country perspective (European Social Survey Round 4 Data, 2008). However, this opportunity also bears the risk of drawing wrong conclusions, if the scales measuring ageism are not comparable across the countries under study. Such comparisons require measurement equivalence across countries. In the current study, utilizing ESS fourth round data from 29 European countries we examine the cross-country measurement equivalence properties of two concepts that are measured by multiple indicators in the module: (1) competence and warmth and (2) experience of age discrimination. We test for measurement equivalence using two analytical methods: multi-group confirmatory factor analysis (MGCFA) and the alignment optimization. Our findings suggest that cross-country comparisons of these measures are trustworthy. We briefly discuss cross-country differences in competence and warmth and experience of age discrimination. Finally, we underline the importance of testing the cross-group equivalence of measurement instruments before using them in different groups (such as countries) so that meaningful substantive conclusions can be drawn.

Keywords: ageism; measurement invariance or measurement equivalence; European Social Survey (ESS); multi-group confirmatory factor analysis (MGCFA); alignment optimization

1 Introduction

Ageism is stereotyping and discrimination against individuals or groups because of their age (e.g., Raymer, Reed, Spiegel, & Purvanova, 2017). Ageism aimed at persons of advanced age is based on the conviction that an individual over a particular age has a lower value and represents a higher economic burden on society, particularly on the younger working population, because of his or her increasing consumption of health and welfare services (e.g., Walker, 1990; Wearing, 1995, 4). Robert Butler introduced the concept of ageism and defined it as a process of systematic stereotyping and discrimination against people based on their old[1] age (Butler, 1969). Although this definition of the concept has become widely accepted (e.g., Iversen, Larsen, & Solem,

Contact information: University of Cologne, Faculty of Management, Economics and Social Sciences, Institute of Sociology and Social Psychology, Albertus-Magnus-Platz, 50923 Cologne, Germany. (E-mail: dseddig@uni-koeln.de)

[1] It should be noted that currently there is no one standard numerical criterion for designating the older population, but most western countries have accepted the chronological age of 60+ years as a definition of 'elderly' or older person. While this definition is somewhat arbitrary, it is frequently associated with the age at which one can begin to receive pension benefits (World Health Organization, 2015). In the present study, we refer to the categorization used in the ESS questionnaire, which defines older people as people over the age of 70.

2009; Kite, Stockdale, Whitley, & Johnson, 2005), further research has shown that ageism can be directed at younger age groups as well (e.g., Ayalon, 2013; Finkelstein, Ryan, & King, 2013; Kite & Johnson, 1988; Raymer et al., 2017). In recent years, a growing body of literature has provided evidence of the existence of ageism aimed at young people (usually defined as people in their 20s and early 30s) – referred to as reverse ageism or adultism (Raymer et al., 2017). Consequently, the concept of ageism was expanded to incorporate prejudice or discrimination against or in favor of any age group (Palmore, 1999). Our study, therefore, is not limited only to attitudes toward older people but also examines the reverse ageism phenomenon.

Although research has shown that ageism prevails across different countries and cultures, its exact intensity in various countries remains to a large extent unknown (e.g., Bodner & Lazar, 2008; Löckenhoff et al., 2009; Xiao, Shen, & Paterson, 2013). Previous studies on attitudes toward different age groups have documented contradictory results regarding differences in the level of ageism across countries. For example, most of the papers reported that attitudes toward older adults tended to be more negative in individualistic countries (e.g., the United States, Australia) than in collectivistic ones (e.g., China, Malaysia, India) (see, e.g., Löckenhoff et al., 2009; Xiao et al., 2013). This finding was in line with the notion that individualistic societies are less concerned with care for their elderly members, which may result in derogative attitudes toward them. Furthermore, individualistic societies place a strong emphasis on performance in the workplace, which is less likely to be attributed to older individuals and may induce such negative attitudes toward them (Hofstede, Hofstede, & Minkov, 2010). However, a number of other studies have shown no significant differences in ageism between individualistic and collectivistic cultures (e.g., Bodner & Lazar, 2008). This contradictory finding may be explained by other mechanisms that may take place in individualistic societies and offset negative attitudes toward older persons. For example, younger people may be perceived in the labor market as a threat for other employees, because they compete for similar positions, whereas older people do not pose such a threat. Yet another possible explanation close at hand for these contrasting findings could be the absence of measurement equivalence across countries, deeming the measures of ageism to be incomparable. Indeed, on the one hand, cross-country differences in ageism may result from differences in response styles or from a different understanding of the concept. On the other hand, country-specific differences in response characteristics or in the understanding of the concept may conceal actual differences. However, with a few notable exceptions (e.g., Bratt, Abrams, Swift, Vauclair, & Marques, 2018; Rupp, Vodanovich, & Credé, 2005), a systematic empirical investigation of the cross-country comparability of the ageism measures remains absent. We argue that cross-country comparability must be examined so that meaningful conclusions about cross-country similarities and differences in ageism may be drawn. Our study, therefore, aims to investigate measurement equivalence of various dimensions of ageism across different European countries. In the current study, we examine three dimensions of ageism (and reverse ageism) drawn from the "Experiences and Expressions of Ageism" module of the ESS (2008/2009) (Abrams, Lima, & Coudin, 2008). The first two dimensions refer to the concepts of competence and warmth of young and old individuals. The third dimension refers to experienced age discrimination.[2] In the following section, we outline the literature that discusses the links between ageism and the three concepts analyzed in this study. Furthermore, we briefly review previous empirical studies that examined ageism from a cross-country perspective. Next, we present the data and variables we employ to measure ageism, explain the methods used to test for measurement equivalence of the scales, and then show the results of our measurement invariance tests. We finalize with a summary and some concluding remarks.

2 Ageism in Previous Studies

Ageism includes two principal components: first, an ageist ideology of stereotypes and attitudes, and second, age discrimination, or behaviors that disadvantage certain persons compared to others based on their age (McMullin & Marshall, 2001; Raymer et al., 2017). In terms of the first component of ageism – ageist stereotypes – a plethora of research highlights the relevance of the Stereotype Content Model (SCM) (e.g., Fiske, Cuddy, Glick, & Xu, 2002; Fiske, Xu, Cuddy, & Glick, 1999; Heckhausen, Dixon, & Baltes, 1989). This model is based on two elements, warmth and competence. It postulates that most stereotypes are mixed, that is, combining positive and negative elements (e.g., Cuddy & Fiske, 2002; Cuddy, Norton, & Fiske, 2005; Fiske et al., 2002). For example, Fiske et al. (2002) demonstrated that elderly are perceived along the competence-warmth spectrum as high in warmth (i.e., socially sensitive and moral) but low in competence (see also Cuddy & Fiske, 2002; Fiske et al., 1999; Heckhausen et al., 1989; Kite, Deaux, & Miele, 1991). Older people are often perceived as less physically attractive, unproductive, depressing, sickly, and even cognitively impaired (e.g., Löckenhoff et al., 2009; Palmore, 1999). Other studies showed that older people have been rated as less ambitious and less responsible (e.g., Andreoletti, Maurice, & Whalen, 2001) compared to younger people. Furthermore, various studies provided evidence for the elderly incompetence stereotype in the labor market, where older workers are considered less effective

[2] In the current study we focused on those concepts in the ESS ageism module that were measured by multiple indicators. We did not examine other items because they did not load on a common latent variable across all countries (see below).

than younger workers (e.g., Avolio & Barrett, 1987; Finkelstein & Farrell, 2007; Fiske et al., 2002; Krings, Sczesny, & Kluge, 2011; Posthuma & Campion, 2009; Singer, 1986). On the other hand, there are also negative stereotypes against young people (e.g., Butler, 1975; Thompson, 1997; Westman, 1991). Some researchers even claimed that an excessive emphasis on the elderly in the ageism literature has revealed that little attention has been paid to younger adults, creating an "ageist ageism literature" (e.g., Rodham, Glover, & Branine, 2001). Studies on reverse ageism, for example, showed that young employees are often perceived as having lower leadership ability (e.g., Collins, Hair Jr, & Rocco, 2009) and as less reliable (e.g., Holt, Marques, & Way, 2012) compared to their older colleagues. This pattern was found inter alia in academia where achievements of the young are often considered as negligible (e.g., Rodham et al., 2001). It should be noted that although there are similarities between ageism toward old and young people with a disadvantaged status in both age groups (Heikkinen & Krekula, 2008), ageism toward old people is essentially different as it is associated with deterioration and death rather than with the developmental potential of youth (Iversen et al., 2009).

In terms of perceived age discrimination – the second dimension of ageism examined in the current study – previous studies have suggested that it may occur in different contexts and perceived by individuals in different age groups. For example, age discrimination based on ageist stereotypes is very common in medical and work environments. It has been shown that medical doctors often provide inferior health care to the elderly (Pasupathi & Löckenhoff, 2002). Moreover, it was found that medical doctors tend to solicit and provide more detailed health information to younger patients (Greene, Adelman, Charon, & Hoffman, 1986; Rost & Frankel, 1993). In the labor market, it is more difficult for older adults to find and keep a job due to stereotypes that older workers have lower job performance and productivity than younger workers (McCann & Giles, 2002). Interestingly enough, previous research shows that young employees report elevated levels of perceived age discrimination as well (e.g., Ayalon, 2013; Kessler, Mickelson, & Williams, 1999). Indeed, age discrimination was found in some cases to not be linear but rather to vary in a curvilinear fashion over the life course (e.g., Gee, Pavalko, & Long, 2007). For example, while examining perceived discrimination in the workplace, Gee et al. (2007) found that reports of age discrimination are relatively high in the 20s, decline in the 30s, and then rise again thereafter. Other research has demonstrated that young employees (under the age of 30) report the highest levels of perceived age discrimination, followed by employees in their 30s, whereas middle-aged (40s) and older workers (50s) report the lowest levels (e.g., Snape & Redman, 2003).

Ageism was examined in previous research across various cultural contexts including the United States (e.g., Fiske et al., 2002; Fiske et al., 1999), Europe (e.g., Ayalon, 2013), and East Asia (e.g., Cuddy et al., 2005; Harwood et al., 1996). These studies found that contrary to expectations, age stereotypes and discrimination extend not only beyond the United States, but also beyond individualistic cultures in general, and are widespread also in collectivistic East Asian countries (e.g., Cuddy et al., 2005). In some cross-country comparative studies, respondents from countries of the Asian Pacific region have demonstrated even more negative attitudes toward the elderly than did their American counterparts (see, e.g., Tien-Hyatt, 1987, for China; Koyano, 1989, for Japan; and Sharps, Price-Sharps, & Hanson, 1998, for Thailand). Cuddy, Fiske, and Glick (2004) examined elderly people and other groups on items measuring warmth and competence, status, and competition in six countries (Belgium, Costa Rica, Hong Kong, Japan, Israel, and South Korea). They found that in all samples, including the East Asian ones, elderly people were viewed as having low status and as noncompetitive. In another study, Israelis held less ageist attitudes compared to respondents in other Western European cultures but similar attitudes to those found in North America, thus providing support for the prevalence and generalizability of ageism in diverse cultures (Bodner & Lazar, 2008). Indeed, there is systematic evidence that ageism is pan-cultural.

Yet only few studies (see, e.g., Bratt et al., 2018) have addressed the question of whether scales measuring attitudes toward the elderly are comparable across countries. As indicated earlier, if they are not, comparisons of scores across countries based on the scale may not be meaningful. The fourth round of the ESS from 2008 provides us with a unique opportunity to measure ageism attitudes and experiences in 29 countries and to test the measurement equivalence characteristics of the items. Thus, in the empirical part we are going to employ these data to examine their measurement equivalence properties across diverse European countries.

3 Data and Variables

For the current analyses, we employ data from the fourth round of the ESS (European Social Survey Round 4 Data, 2008) conducted from 2008 to 2009. This particular round of the ESS contains data related to age and ageism, which have been collected in the context of the "Experiences and Expressions of Ageism" module (Abrams et al., 2008). Data are available from 29 European countries with a total of 56,752 respondents aged 15 or older who reside in private households, regardless of their nationality, citizenship, language used or legal status in the country (European Social Survey, 2010). The participating countries are Belgium (1,760), Bulgaria (2,230), Croatia (1,484), Cyprus (1,215), Czech Republic (2,018), Denmark (1,610), Estonia (1,661), Finland (2,195), France (2,073), Germany (2,751), Greece (2,072), Hungary (1,544), Ireland (1,764), Israel (2,490),

Latvia (1,980), Netherlands (1,778), Norway (1,549), Poland (1,619), Portugal (2,367), Romania (2,146), Russia (2,512), Slovenia (1,286), Slovakia (1,810), Spain (2,576), Sweden (1,830), Switzerland (1,819), Turkey (2,416), Ukraine (1,845), and United Kingdom (2,352) (the number of respondents per country is given in parentheses). Since 150 persons had missing values on all variables used in our study, the total number of respondents in our study reduced to 56,602. Further information on data collection and documentation is available on the ESS website: www.europeansocialsurvey.org.

The "Experiences and Expressions of Ageism" module contains various major concepts and subdimensions.[3] The current study focuses on concepts measured by multiple items that can be analyzed as latent factors. These are competence and warmth, and the experience of discrimination.[4] The concept of competence and warmth targets different age groups, that is, people in their 20s and people over 70, whereas the concept of experienced discrimination due to age targets the respective age group of the respondent. The items measuring these dimensions and the response scales are presented in Table 1.

The questions measuring competence and warmth inquire – separately for people in their 20s and over 70 – whether they are friendly, competent, having high moral standards and respected. The response categories ranged between 0 (*not at all likely*) and 4 (*very likely*). The questions measuring experienced age discrimination inquired how often in the past one has been prejudiced or treated unfairly, felt lack of respect, and been treated badly because of the respondent's age. The response categories ranged between 0 (*never*) and 4 (*very often*).

4 Method

Cross-group equivalence of measurements of a latent construct (Bollen, 1989) is essential in order to draw meaningful conclusions in cross-cultural or longitudinal research. Absence of measurement equivalence is problematic because cross-country differences in mean scores or in regression coefficients across groups may be a result of systematic biases in responses across groups or due to a different understanding of the questions rather than due to real differences. On the other hand, finding no differences in mean scores or regression coefficients across groups does not mean that "real" differences are absent (Davidov et al., 2015; Davidov et al., 2018, 4; Davidov, Meuleman, Cieciuch, Schmidt, & Billiet, 2014; Horn & McArdle, 1992; Steenkamp & Baumgartner, 1998, 1; Vandenberg & Lance, 2000). Thus, it is necessary to test for measurement invariance and determine "whether or not, under different conditions of observing and studying phenomena, measurement operations yield measures of the same attribute" (Horn & McArdle, 1992, p. 117). Among the many methods to assess measurement invariance, multiple-group confirmatory factor analysis (MGCFA) is one of the most commonly used ones (Jöreskog, 1971; Reise, Widaman, & Pugh, 1993), as it allows testing for different hierarchical levels of measurement invariance (Horn & McArdle, 1992; Meredith, 1993; Vandenberg & Lance, 2000).

Configural invariance refers to equivalence of the measurement structure. This means that each factor has the same set of indicators in each group, the model fits the data well in each group, all factor loadings are substantial, and correlations among factors are less than one. This does not imply, however, that the meaning of a latent variable is the same across groups. *Metric* invariance additionally requires factor loadings to be equal across groups and is deemed necessary to compare factor covariances and unstandardized regression coefficients. Finally, scalar measurement invariance additionally requires the intercepts of the indicators to be equal across groups. When the analyses demonstrate scalar measurement invariance, we can confidently compare the means of the latent variables across groups.

We considered three global fit measures: the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardized root mean residual (SRMR). We considered models with a CFI value higher than 0.90, and RMSEA and SRMR values lower than 0.08 as acceptable (see Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004; West, Taylor, & Wu, 2012). We adopted a "bottom-up" analytical strategy that begins with the least restrictive (configural) model. We then consecutively introduce restrictions on the factor loadings (metric model) and intercepts (scalar model). If the model fit does not significantly decrease by imposing additional restrictions, the more restrictive model can be accepted. However, traditional difference tests based on chi-square values are sensitive to sample size (Saris, Satorra, & Sörbom, 1987). Therefore, we follow the guidelines suggested by Chen (2007) who recommends taking the differences in other (chi-square based) model fit statistics into consideration, that is, the CFI, RMSEA, and SRMR. With $N > 300$, differences between configural and metric models are considered relevant when the change in CFI is larger than 0.010, accompanied by a change in RMSEA larger than 0.015, or a change in SRMR larger than 0.030. Differences

[3] A full description of the concepts and items used in the module can be found on the ESS website (http://www.europeansocialsurvey.org/docs/round4/questionnaire/ESS4_final_ageism_module_template.pdf).

[4] We use the terms latent dimension and latent factor interchangeably. We do not consider the concepts benevolent and hostile forms of prejudice and perceived threat which are also covered by the module by multiple indicators, because preliminary analyses have demonstrated that the single measures are rather independent and do not load on a common latent variable. We re-estimated all models also while excluding people in their 20s and/or over 70. All results remained essentially the same. These preliminary analyses may be obtained from the first author upon request.

Table 1
Items measuring competence and warmth and experience of discrimination

Item	Question wording	Answers
Competence and warmth of people in their 20s (cw20)		
1	Most people view those in their 20s as friendly	Not at all likely (0) – Very likely (4)
2	Most people view those in their 20s as competent	Not at all likely (0) – Very likely (4)
3	Most people view those in their 20s as having high moral standards	Not at all likely (0) – Very likely (4)
4	Most people view those in their 20s with respect	Not at all likely (0) – Very likely (4)
Competence and warmth of people over 70 (cw70)		
5	Most people view those over 70 as friendly	Not at all likely (0) – Very likely (4)
6	Most people view those over 70 as competent	Not at all likely (0) – Very likely (4)
7	Most people view those over 70 as having high moral standards	Not at all likely (0) – Very likely (4)
8	Most people view those over 70 with respect	Not at all likely (0) – Very likely (4)
Experience of age discrimination (exp_age_dis)		
9	. . . please tell me how often, in the past year, anyone has shown prejudice against you or treated you unfairly because of your age	Never (0) – Very often (4)
10	. . . 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 patronising you?	Never (0) – Very often (4)
11	. . . 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?	Never (0) – Very often (4)

between metric and scalar invariance are considered relevant when the change in CFI is larger than 0.01, the change in RMSEA is larger than 0.015, and the change in SRMR is larger than 0.01.

Recent literature suggests that MGCFA may be too strict. In other words, it may suggest that measurement invariance is not given, even though mean comparisons may be meaningful. Therefore, Asparouhov and Muthén (2014, 4) recently proposed a more flexible method to test for measurement invariance that does not impose such strict restrictions on the measurement model, the alignment optimization procedure (see also Marsh et al., 2018, 3; B. O. Muthén & Asparouhov, 2014, 2018, 3; for an application, see Cieciuch, Davidov, Algesheimer, & Schmidt, 2018, 4; Cieciuch, Davidov, & Schmidt, 2018; Munck, Barber, & Torney-Purta, 2018, 4). The alignment method is more lenient, that is, it does not require exact equality of parameters across groups. The procedure begins with an unrestricted (configural) base model in which the factor loadings and intercepts are freely estimated without cross-group equality constraints and the factor means are fixed to zero. This is the best fitting model. The final alignment model has the same fit as the base model. However, now the factor means are estimated so that the degree of noninvariance of the aligned parameters is kept to a minimum. This is obtained by minimizing a loss (simplicity) function similar to rotation in exploratory factor analysis (Asparouhov & Muthén, 2014, 4). That is, compared to the base model, the final model will yield as many approximately invariant parameters as possible and only a few large noninvariant parameters. In the FREE alignment procedure, the latent variance of a reference group is fixed to 1 and the factor mean of that group is estimated. In the FIXED alignment procedure, the factor mean of the reference group is fixed to 0. The choice of a reference group is not arbitrary, and it is usually the group with mean closest to zero (Asparouhov & Muthén, 2014, 4). A criterion to assess the acceptability and trustworthiness of an alignment solution is the proportion of noninvariant parameters. B. O. Muthén and Asparouhov (2014) recommend that less than 25% of the parameters should be noninvariant. Flake and McCoach (2018) refer to the situation where researchers are particularly interested in factor mean comparisons. They suggest inspecting the extent of noninvariance of factor loadings and intercepts separately and suggest that less than 29% of the intercepts should be noninvariant. In case these cut-off values are exceeded, the trustworthiness of the estimated factor means can be assessed with a Monte Carlo simulation study. In a Monte Carlo simulation study, hypothesized population parameter values are used to generate a large number of artificial data samples (Paxton, Curran, Bollen, Kirby, & Chen, 2001). Subsequently, a hypothesized model is estimated for each sample, which makes it possible to assess the precision of the replicated parameter estimates, standard errors, and tests of model fit. In this study, the estimates of the final alignment models are used for the data generation, and the correlation between the generated and estimated factor means is monitored to assess the quality of the arrangements of the estimated factor means (Asparouhov & Muthén, 2014, 4; B. O.

Muthén & Asparouhov, 2014). Thus, if the simulation reveals a very high degree of parameter recovery, factor means may be comparable, even if a large degree of noninvariance exists in the alignment model. This is achieved when correlations between generated and estimated factor means are equal to or higher than 0.98. One may then conclude that measurement invariance is given based on the alignment approach, even when the MGCFA approach suggests that measurement invariance is not supported by the data.

As discussed below, specific item intercepts displayed large variations across countries whereas the cross-country differences of intercepts of other items were more moderate. The alignment procedure is particularly equipped to examine measurement invariance in such situations, and it is also very comfortable to use when the number of groups is large, as in our case. Thus, by applying the alignment method, we tested whether the pattern of a few large noninvariant parameters and many more similar parameters holds in the present data. Since alignment yields an exploratory solution, it can be used as a starting point for more informed multiple-group CFA models that test for partial MI (Byrne, Shavelson, & Muthén, 1989) or models that assess alternative forms of approximate MI using the Bayesian approach (Cieciuch, Davidov, Algesheimer, & Schmidt, 2018, 4; Cieciuch, Davidov, Schmidt, Algesheimer, & Schwartz, 2014; Lek et al., n.d.; B. O. Muthén & Asparouhov, 2013; Seddig & Leitgöb, 2018a, 2018b; Van de Schoot et al., 2013). However, we do not consider these other approaches in this study.

We used the software package M*plus* Version 8 (L. Muthén & Muthén, 2017) for all calculations. We treated the five-point rating scale as continuous and used maximum-likelihood estimation with robust adjustment for standard errors (Rhemtulla, Brosseau-Liard, & Savalei, 2012) to deal with the problem of missing values (Schafer & Graham, 2002).[5] Two items measuring experience of age discrimination (item 9 measuring experience of prejudice and item 11 measuring being treated badly because of age) were skewed, and most of the responses appeared in the lowest category ("never"). When we treated these items as categorical instead and used weighted least squares estimation (the WLSMV estimator in M*plus*: see, e.g., Beauducel & Herzberg, 2006; Flora & Curran, 2004; B. O. Muthén, du Toit, & Spisic, 1997), invariance tests produced similar results to those obtained when using maximum-likelihood while considering the items to be continuous and normally distributed.[6]

5 Results

5.1 Confirmatory Factor Analysis

First, we used confirmatory factor analysis (CFA) to test a model with three latent variables for the entire ESS sample: competence and warmth of people in their 20s, competence and warmth of people over 70, and experience of discrimination. The items inquiring about competence and warmth of young and older people did not allow distinguishing between competence and warmth as separate categories, because items measuring warmth and competence loaded on a single latent variable for each age group. Following Vauclair, Abrams, and Bratt (2010), these items operated as a "superordinate stereotype index". Thus, we used a single factor for competence and warmth for each age group respectively in all countries. For the third factor we used the three items measuring experience of discrimination. The model is presented in Figure 1. The model fit the data well (CFI = 0.983, RMSEA = 0.034, SRMR = 0.019), and all items loaded strongly on their target latent variables with standardized factor loadings ranging between 0.583 and 0.906.[7] Next, we ran this model in each country separately. The model also attained acceptable fit in each country. The standardized factor loadings of the items ranged between 0.42 and 0.96, indicating that the model structure applied in all countries (Brown, 2015).[8] We used this model for the subsequent cross-country measurement invariance tests.

Competence and warmth ascribed to people in their 20s correlated positively with competence and warmth toward people over 70 (see Figure 1). Furthermore, both dimensions correlated negatively with experience of discrimination. The same pattern emerged when the model was analyzed per country. The highest correlations between competence and warmth ascribed to people in their 20s and over 70 were found in Turkey (r = .68), Cyprus (r = .65), and Croatia (r = .62). The negative correlations between competence and warmth ascribed to people in their 20s and experience of discrimination were strongest in Greece (r = −.23), Latvia (r = −.22), and Turkey (r = −.22). The strongest negative correlations between competence and warmth toward people over 70 and experience of discrimination were found in Turkey (r = −.28), Israel (r = −.25), and Greece (r = −.18). However, in some countries (Cyprus, Estonia, France, Ireland, Netherlands, Poland, Slovenia, Spain, and Ukraine), single correlations between warmth and competence constructs and experience of age discrimination were not significant.

[5] On average, only 4% of the responses were missing.

[6] All M*plus* codes are available on the journal's website. A more detailed output is available from the first author upon request.

[7] We considered allowing error correlations between the items measuring warmth and competence for different age groups. Although the model fit improved, the correlations were low (mostly below 0.1 or even negative). Therefore, we opted for the simpler model without error correlations.

[8] There was only one exception – Poland. In this country, the factor loading for the item measuring the perceived competence of people over 70 was only 0.346. The output for the CFA models in specific countries may be obtained from the first author upon request.

5.2 Testing for measurement invariance using MGCFA

We then turned to the model testing for configural invariance across 29 countries. The configural model fit the data well, as demonstrated in Table 2. We introduced equality constraints on the factor loadings across countries to test for metric measurement invariance across the 29 countries. The model fit was acceptable and not substantially worse than the fit of the configural model. Although the decline of the CFA was above the cutoff value suggested by Chen (2007), the decline in RMSEA and SRMR was not. Thus, we could conclude that metric invariance was supported by the data. This implied that comparisons of factor covariance and unstandardized regression estimates for the three latent variables across the 29 countries were now possible. To test for scalar invariance, we added cross-country equality constraints on the intercepts. Model fit indices indicated a substantial deterioration in the fit of the scalar invariance model (Chen, 2007). Thus, based on this test, cross-country comparisons of factor mean scores of the three latent variables may not be performed with confidence. Variation in intercepts was particularly evident for the items 2 and 4 (viewing people in their 20s as competent and with respect), 5 and 6 (viewing people over 70 as friendly and competent), and 9 (being prejudiced unfairly because of age), whereas cross-country differences in the intercepts of other items were somewhat more moderate.

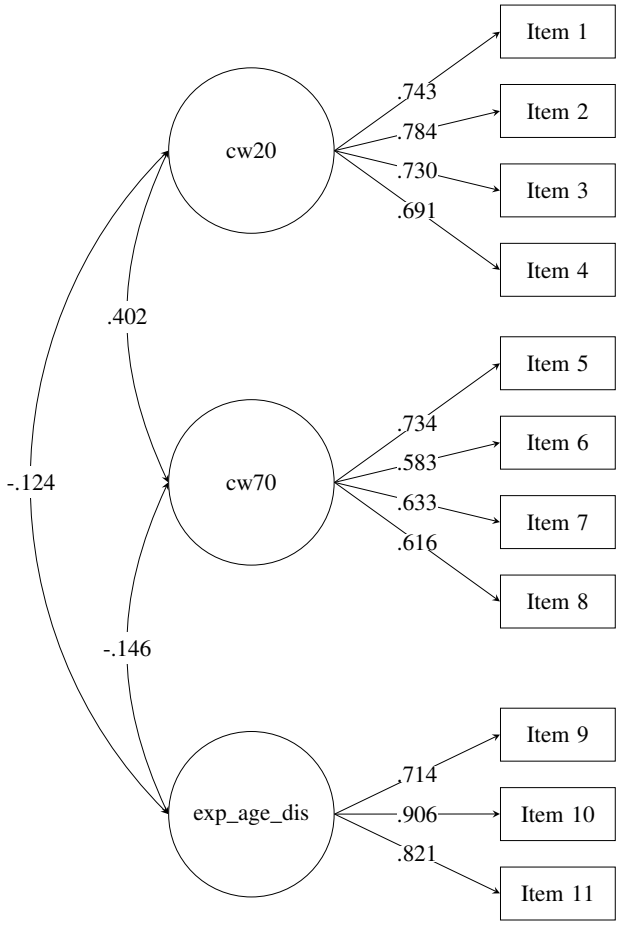

Figure 1. Baseline confirmatory factor model estimated on the total sample (N = 56,602). Note: All factor loadings and covariances between the latent variables are standardized and statistically significant for the total sample ($p < 0.000$). For latent variable and item labels, see Table 1.

5.3 Testing for measurement invariance using the alignment optimization

In the next step, we assessed the measurement invariance properties of the three scales in the ESS based on the alignment approach. Since this approach is not as strict as the MGCFA, it could well be the case that it would suggest that measurement invariance is given, even though the stricter MGCFA suggests the opposite (Asparouhov & Muthén, 2014, 4; Marsh et al., 2018, 3; B. O. Muthén & Asparouhov, 2014). We estimated the alignment models separately for each latent dimension. The results suggested to use the FIXED optimization procedure to obtain unbiased parameters and standard errors.

As shown in Table 3, the proportion of noninvariant factor loadings for each latent dimension was low and considerably below 25%. However, the proportion of noninvariant intercept parameters for each latent dimension exceeded the recommended 29% cut-off value. Thus, it suggested that factor means may not be comparable. Bratt et al. (2018) reached similar conclusions when they examined the measurement equivalence properties of the perceived age discrimination concept. To circumvent this problem, the authors suggested using results from the alignment analysis to cluster the countries into two groups, each with a high degree of measurement invariance. Approximate measurement invariance (approximately invariant factor loadings and approximately invariant thresholds) was satisfied using their approach, with few noninvariant parameters for single countries. We decided to use an alternative approach to examine whether all 29 countries display measurement noninvariance for the three concepts. Instead of grouping the countries, we investigated whether the arrangement of means is trustworthy in spite of the evidenced nonequal parameters across countries.

To do this, we conducted several Monte Carlo simulations. We used the final estimates of the alignment models as starting values for the data generation process and checked the replication of the factor means (see Asparouhov & Muthén, 2014, 4; Marsh et al., 2018, 3; B. O. Muthén & Asparouhov, 2014).

The correlation between the generated and estimated factor means was monitored for four sample sizes per group: 100, 500, 1,000, and 2,000. The latter condition refers to

Table 2
Traditional MGCFA model fit across 29 countries (N = 56,602)

Model	χ^2	(df)	CFI	RMSEA	SRMR
Configural	5925.227	(1131)	0.969	0.047	0.034
Metric	8520.479	(1355)	0.954	0.052	0.052
Scalar	23688.098	(1579)	0.858	0.085	0.078

Note: χ^2 = chi-square, df = degrees of freedom, CFI = comparative fit index, RMSEA = root mean square error of approximation, SRMR = standardized root mean residual.

Table 3
Noninvariance in the alignment MGCFA across 29 countries

	Factor loadings		Intercepts	
	Noninvariant (total)	Noninvariant (%)	Noninvariant (total)	Noninvariant (%)
Competence & warmth 20	28 (116)	24.14	50 (116)	43.10
Competence & warmth 70	19 (116)	16.38	76 (116)	65.52
Experience of age discrimination	19 (87)	21.84	46 (87)	52.87

Note: In columns 2 and 4, the first number refers to the number of noninvariant factor loadings and intercepts, respectively. The second number (in parentheses) refers to the total number of factor loadings and intercepts, respectively. For example, looking at the second row, 28 out of 116 factor loadings (24.14%) of cw20 are noninvariant.

a situation that is similar to our real data with some 58,000 observations in total across 29 countries. Each simulation included 500 replications. The results of the simulation are presented in Table 4. With small sample sizes they indicated that the factor means would not be trustworthy. However, with sample sizes similar to those of the ESS of 1,000 and 2,000 per country, all correlations between the generated and the estimated factor means were large enough and exceeded 0.98, implying that the factor means, based on the alignment procedure, were trustworthy after all. Table 5 lists the factor means for the three latent dimensions based on the alignment solutions.

The estimated factor means revealed some interesting cross-country differences. People in their 20s were most strongly perceived to be competent and warm in Israel, followed by Cyprus, Greece, Turkey, Slovenia, Spain, and Portugal. The lowest mean was estimated in the Ukraine. Also several other eastern European countries (e.g., Estonia, Bulgaria, and Russia) and the UK displayed low mean scores for this latent variable. People over 70 were most strongly perceived as competent and warm in Hungary and Ireland. The lowest mean was estimated in several eastern European countries (e.g., Romania, Croatia, Czech Republic, and Slovakia). Perceived age-related discrimination was highest in east European countries (e.g., Czech Republic, Romania, Russia, Slovakia, and Ukraine). The lowest level of perceived age discrimination existed in a mixture of countries from southern, northern, and central Europe (e.g., Portugal, Cyprus, Denmark, Norway, and Switzerland).

To explore country mean differences in a more systematic way, we examined how these differences related to the level of individualism and collectivism in these societies. As indicated previously, we expect that individualistic societies would display higher scores of ageism toward older individuals. In such societies, one is less concerned with care for elderly members, which may result in derogating attitudes toward them. Furthermore, individualistic societies place a strong emphasis on performance in the workplace, which is less likely to be attributed to older individuals and may induce negative attitudes toward them (Hofstede et al., 2010). The relationships between the degree of individualism in each country participating in our ESS sample and its respective factor mean score for the three latent variables are plotted in Figure 2.[9] In contrast to our expectations, attitudes were more favorable toward older people and less favorable toward younger people in individualistic societies, as evident in the negative correlation of competence and warmth (20s) and the positive correlation of competence and warmth (over 70) with individualism. It could well be the case that particularly in individualistic societies, which are often characterized by stronger competition in various life domains such as the labor market, younger people may be perceived as a threat because they compete for similar resources, whereas

[9]The individualism scores were obtained from https://www.hofstede-insights.com/country-comparison/. Higher scores indicated a higher level of individualism in a country. No data was available for Cyprus; therefore, it was excluded from this analysis.

Table 4
Alignment simulation: Correlations between generated and estimated factor means across 29 groups (500 replications)

	N_g=100	N_g=500	N_g=1,000	N_g=2,000
Competence & warmth 20	0.9474	0.9869	0.9947	0.9974
Competence & warmth 70	0.8881	0.9754	0.9877	0.9937
Experience of age discrimination	0.8798	0.9732	0.9845	0.9930

older people do not pose such a threat. As a result, negative attitudes toward younger individuals may be more prevalent in individualistic societies. However, strikingly, people in individualistic societies perceived age-related discrimination to a lower extent compared to their counterparts in collectivist societies, as evidenced by the negative correlation between individualism and the perceived age discrimination index.

6 Summary and Discussion

Ageism is stereotyping and discrimination against individuals or groups because of their age (e.g., Raymer et al., 2017). Indeed, the concept of ageism incorporates prejudice or discrimination against or in favor of any age group. Whereas previous studies on attitudes toward different age groups have explored and documented differences in the levels of ageism across countries, only few actually examined whether these attitudes were comparable across countries. Testing for measurement invariance of the scale prior to comparing its scores across countries is a prerequisite, however, to allow researchers to draw valid conclusions about similarities and differences. After all, cross-country differences in ageism may result from differences in response styles or from a different understanding of the concept. At the same time, country-specific differences in response characteristics or in the understanding of the concept may conceal actual differences. Thus, in the current study we aimed to fill this gap and investigate measurement equivalence of dimensions of ageism across different European countries. We utilized data from the fourth round of the ESS, which included a rotating module on the topic of ageism with multiple item measures of three concepts: perceptions of competence and warmth of people in their 20s, similar perceptions of people over the age of 70, and perceived (own) age discrimination. The analyses were performed using two analytical methods: MGCFA and alignment optimization.

While the results of the MGCFA analysis revealed lack of support for measurement invariance of the scales, the findings based on the more lenient alignment procedure were much more promising and suggested that measurement equivalence of the scales was given in the data. Thus, the findings implied that the ESS ageism scales were comparable after all across all the countries participating in the ESS survey, and their scores may be compared across countries with confidence. We revealed considerable differences both in the level of prejudice toward people in their 20s and toward people over 70 and in the scores of perceived age discrimination and attempted to explain or characterize these differences by observing their patterns in countries that are more or less individualistic.

This study is not without limitations. First, the present study tested for measurement invariance of ageism questions on a group of respondents with different ages that represent the general population. We analyzed the answers of all respondents available in the data. However, this does not necessarily mean that every age group within the general population interprets the items in the same way across countries. Future research may try to assess the measurement invariance properties of the ageism questions across respondents belonging to different age groups, to examine whether people of a different age respond to and understand questions on ageism in a similar way. Second, the module included several other questions measuring different aspects of ageism. However, we could not examine their cross-country measurement invariance properties, because they did not display a consistent measurement pattern across countries or did not load on their respective latent variables. Some of these items measured specific and unique aspects rather than the content they were meant to measure. Therefore, we had to exclude them from further analysis and focus on those multiple item measures which displayed high reliability across all countries. Future studies may try to improve the measurement reliability of these excluded items so that they load more strongly on their respective latent variables. Third, competence and warmth should ideally be measured by two separate sets of questions to reflect two distinct dimensions of prejudice. However, these items, as measured in the ESS, correlated too strongly so that it was not possible to separate them (Vauclair et al., 2010). Consequently, we had to unify them into a single construct measuring both competence and warmth for each age group, respectively. This unified construct represented a general tendency to express prejudice. Future studies should design better measures of warmth and competence that allow discriminating between the two dimensions of warmth and competence according to theoretical expectations. Notwithstanding these limitations, the current study was one of the first to assess the cross-country measurement comparability of *various* measures of ageism

Table 5
Alignment based factor mean comparison across countries

	Competence & warmth 20 (N = 55,440; reference group: Israel)		Competence & warmth 70 (N = 55,595; reference group: Spain)		Experience of age discrimination (N = 56,469; reference group: Portugal)	
Rank	Country	Mean	Country	Mean	Country	Mean
1	Israel	0.000	Hungary	0.300	Czech Rep.	1.246
2	Cyprus	−0.211	Ireland	0.241	Romania	0.949
3	Greece	−0.239	Spain	0.000	Russia	0.898
4	Turkey	−0.289	Greece	−0.081	Slovakia	0.871
5	Slovenia	−0.322	Israel	−0.098	Ukraine	0.836
6	Spain	−0.399	Norway	−0.105	Turkey	0.730
7	Portugal	−0.454	Portugal	−0.114	Israel	0.667
8	Latvia	−0.519	Finland	−0.151	France	0.610
9	Poland	−0.540	Sweden	−0.196	Bulgaria	0.579
10	Finland	−0.561	Cyprus	−0.237	Finland	0.543
11	Ireland	−0.569	Turkey	−0.273	UK	0.519
12	Croatia	−0.590	Netherlands	−0.292	Slovenia	0.496
13	Czech Rep.	−0.599	Poland	−0.318	Belgium	0.496
14	France	−0.667	Latvia	−0.336	Ireland	0.481
15	Sweden	−0.717	Ukraine	−0.338	Greece	0.480
16	Denmark	−0.740	UK	−0.357	Germany	0.468
17	Norway	−0.790	Slovenia	−0.357	Hungary	0.457
18	Belgium	−0.842	Denmark	−0.382	Estonia	0.455
19	Romania	−0.855	Bulgaria	−0.384	Poland	0.451
20	Slovakia	−0.982	Estonia	−0.414	Sweden	0.442
21	Switzerland	−0.995	Belgium	−0.547	Spain	0.413
22	Netherlands	−1.047	Russia	−0.557	Latvia	0.400
23	Hungary	−1.057	France	−0.594	Croatia	0.398
24	Estonia	−1.130	Germany	−0.606	Netherlands	0.355
25	Germany	−1.141	Romania	−0.710	Switzerland	0.333
26	Bulgaria	−1.154	Croatia	−0.744	Norway	0.297
27	UK	−1.333	Switzerland	−0.805	Denmark	0.264
28	Russia	−1.379	Czech Rep.	−0.810	Cyprus	0.173
29	Ukraine	−1.402	Slovakia	−0.871	Portugal	0.000

across many countries in a systematic way and using state of the art methods.

In sum, ageism is a phenomenon that is evidently not unique to certain contexts, but is rather widespread, present in more or less individualistic societies, and in different European countries. The ESS provides a unique opportunity to study ageism and explain it throughout Europe. Findings of cross-country measurement invariance of the concepts of ageism that we investigated in the current study allow researchers to draw conclusions about differences and similarities in ageism and its determinants with confidence. Researchers interested in comparing the prevalence and extent of ageism across other groups such as geographical regions in Europe, language groups, age groups within countries, or cultures, should apply procedures similar to the ones presented and implemented in the current study to examine if their concepts of interest display a sufficient level of equivalence across these groups. We hope that the current study will thus help researchers in their endeavor to conduct a meaningful comparative study of ageism.

Acknowledgements

The third author would like to thank the University of Zurich Research Priority Programme "Social Networks" for their support. The authors would like to thank Lisa Trierweiler for the English proof of the manuscript.

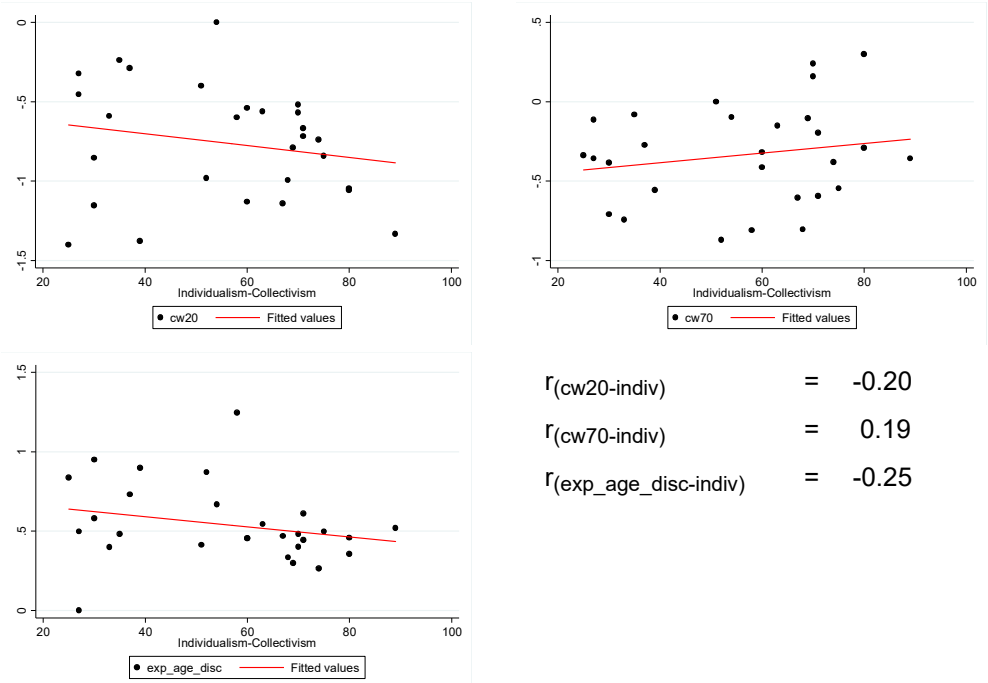

Figure 2. Country-level correlations between the individualism-collectivism scale and the three latent variables of ageism. For latent variable abbreviations, see Table 1.

References

Abrams, D., Lima, L., & Coudin, G. (2008). Experiences and expressions of ageism. A proposal for a rotating module for the 4th round of the European Social Survey. Retrieved from http://www.europeansocialsurvey.org/docs/round4/questionnaire/ESS4_abrams_proposal.pdf

Andreoletti, C., Maurice, J. K., & Whalen, H. (2001). Gender, race, and age: Compound stereotypes across the lifespan. Poster presented at the Second Annual Meeting of the Society for Personality and Social Psychology, San Antonio, TX.

Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling*, *21*, 495–508.

Avolio, B. J., & Barrett, G. V. (1987). Effects of age stereotyping in a simulated interview. *Psychology and Aging*, *2(1)*, 56–63.

Ayalon, L. (2013). Feelings towards older vs. younger adults: Results from the European Social Survey. *Educational Gerontology*, *39(12)*, 888–901.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, *13(2)*, 186–203.

Bodner, E., & Lazar, A. (2008). Ageism among Israeli students: Structure and demographic influences. *International Psychogeriatrics*, *20*(5), 1046–1058.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Bratt, C., Abrams, D., Swift, H. J., Vauclair, C.-M., & Marques, S. (2018). Perceived age discrimination across age in Europe: From an ageing society to a society for all ages. *Developmental Psychology*, *54(1)*, 167–180.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd). New York, NY: Guilford Press.

Butler, R. N. (1969). Age-ism: Another form of bigotry. *The Gerontologist*, *9(4/1)*, 243–246.

Butler, R. N. (1975). Psychiatry and the elderly: An overview. *The American Journal of Psychiatry*, *132(9)*, 893–900.

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.

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14(3)*, 464–504.

Cieciuch, J., Davidov, E., Algesheimer, R., & Schmidt, P. (2018). Testing for approximate measurement invariance of human values in the European Social Sur-

vey. *Sociological Methods & Research*, *47*, 665–686. doi:10.1177/0049124117701478

Cieciuch, J., Davidov, E., & Schmidt, P. (2018). Alignment optimization: Estimation of the most trustworthy means in cross-cultural studies even in the presence of noninvariance. In E. Davidov, P. Schmidt, J. Billiet, & B. Meuleman (Eds.), *Cross-cultural analysis: Methods and applications* (pp. 571–593). New York: Routledge.

Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., & Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: A cross-country illustration with a scale to measure 19 human values. *Frontiers in Psychology*, *5*, 1–10.

Collins, M. H., Hair Jr, J. F., & Rocco, T. S. (2009). The older-worker-younger-supervisor dyad: A test of the reverse Pygmalion effect. *Human Resource Development Quarterly*, *20(1)*, 21–41.

Cuddy, A. J., & Fiske, S. T. (2002). Doddering, but dear: Process, content, and function in stereotyping of elderly people. In T. D. Nelson (Ed.), *Ageism* (pp. 3–26). Cambridge, MA: MIT Press.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2004). When professionals become mothers, warmth doesn't cut the ice. *Journal of Social Issues*, *60(4)*, 701–718.

Cuddy, A. J., Norton, M. I., & Fiske, S. T. (2005). This old stereotype: The pervasiveness and persistence of the elderly stereotype. *Journal of Social Issues*, *61(2)*, 267–285.

Davidov, E., Cieciuch, J., Meuleman, B., Schmidt, P., Algesheimer, R., & Hausherr, M. (2015). The comparability of measurements of attitudes toward immigration in the European Social Survey: Exact versus approximate measurement equivalence. *Public Opinion Quarterly*, *79(S1)*, 244–266.

Davidov, E., Dülmer, H., Cieciuch, J., Kuntz, A., Seddig, D., & Schmidt, P. (2018). Explaining measurement nonequivalence using multilevel structural equation modeling: The case of attitudes toward citizenship rights. *Sociological Methods & Research*, *47*, 729–760. doi:10.1177/0049124116672678

Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, *40*, 55–75.

European Social Survey. (2010). ESS round 4 (2008/2009) final activity report. London: Centre for Comparative Social Surveys, City University London.

European Social Survey Round 4 Data. (2008). Data file edition 4.4. NSD - Norwegian Centre for Research Data, Norway – Data Archive and distributor of ESS data for ESS ERIC.

Finkelstein, L. M., & Farrell, S. K. (2007). An expanded view of age bias in the workplace. In K. S. Shultz & G. A. Adams (Eds.), *Applied psychology series. Aging and work in the 21st century* (pp. 73–108). Mahwah, NJ: Lawrence Erlbaum Associates.

Finkelstein, L. M., Ryan, K. M., & King, E. B. (2013). What do the young (old) people think of me? Content and accuracy of age-based metastereotypes. *European Journal of Work and Organizational Psychology*, *22(6)*, 633–657.

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82(6)*, 878–902.

Fiske, S. T., Xu, J., Cuddy, A. C., & Glick, P. (1999). (dis) respecting versus (dis) liking: Status and interdependence predict ambivalent stereotypes of competence and warmth. *Journal of Social Issues*, *55(3)*, 473–489.

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*, *25(1)*, 56–70.

Flora, B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, *9(4)*, 466–491.

Gee, G. C., Pavalko, E. K., & Long, J. S. (2007). Age, cohort and perceived age discrimination: Using the life course to assess self-reported age discrimination. *Social Forces*, *86(1)*, 265–290.

Greene, M. G., Adelman, R., Charon, R., & Hoffman, S. (1986). Ageism in the medical encounter: An exploratory study of the doctor-elderly patient relationship. *Language & Communication*, *6(1)*, 113–124.

Harwood, J., Giles, H., Ota, H., Pierson, H. D., Gallois, C., Ng, S. H., & Somera, L. (1996). College students' trait ratings of three age groups around the Pacific Rim. *Journal of Cross-Cultural Gerontology*, *11(4)*, 307–317.

Heckhausen, J., Dixon, R. A., & Baltes, P. B. (1989). Gains and losses in development throughout adulthood as perceived by different adult age groups. *Developmental Psychology*, *25(1)*, 109–121.

Heikkinen, S., & Krekula, C. (2008). lderism—ett fruktbart begrepp? [Ageism—a useful concept?] *Sociologisk Forskning*, *45(2)*, 18–34.

Hofstede, G., Hofstede, G. J., & Minkov, M. (2010). *Cultures and organizations: Software of the mind* (3rd). New York, NY: McGraw-Hill.

Holt, S., Marques, J., & Way, D. (2012). Bracing for the millennial workforce: Looking for ways to inspire gen-

eration Y. *Journal of Leadership, Accountability and Ethics*, *9(6)*, 81–93.

Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18(3)*, 117–144.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6(1)*, 1–55.

Iversen, T. N., Larsen, L., & Solem, P. E. (2009). A conceptual analysis of ageism. *Nordic Psychology*, *61(3)*, 4–22.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36(4)*, 409–426.

Kessler, R. C., Mickelson, K. D., & Williams, D. R. (1999). The prevalence, distribution, and mental health correlates of perceived discrimination in the United States. *Journal of Health and Social Behavior*, *40(3)*, 208–230.

Kite, M. E., Deaux, K., & Miele, M. (1991). Stereotypes of young and old: Does age outweigh gender? *Psychology and Aging*, *6(1)*, 19–27.

Kite, M. E., & Johnson, B. T. (1988). Attitudes toward older and younger adults: A meta-analysis. *Psychology and Aging*, *3(3)*, 233–244.

Kite, M. E., Stockdale, G. D., Whitley, B. E., & Johnson, B. T. (2005). Attitudes toward younger and older adults: An updated meta-analytic review. *Journal of Social Issues*, *61(2)*, 241–266.

Koyano, W. (1989). Japanese attitudes toward the elderly: A review of research findings. *Journal of Cross-Cultural Gerontology*, *4(4)*, 335–345.

Krings, F., Sczesny, S., & Kluge, A. (2011). Stereotypical inferences as mediators of age discrimination: The role of competence and warmth. *British Journal of Management*, *22(2)*, 187–201.

Lek, K., Oberski, D., Davidov, E., Cieciuch, J., Seddig, D., & Schmidt, P. (n.d.). Approximate measurement invariance. In T. Johnson, B. Pennell, I. Stoop, & B. Dorer (Eds.), *Advances in comparative survey methods: Multinational, multiregional and multicultural contexts (3MC)*. in press. New York: Wiley.

Löckenhoff, C. E., De Fruyt, F., Terracciano, A., McCrae, R. R., De Bolle, M., Costa Jr, P. T., & Allik, J. (2009). Perceptions of aging across 26 cultures and their culture-level associates. *Psychology and Aging*, *24(4)*, 941–954.

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*, 524–545. doi:10.1037/met0000113

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). Search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers of overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11(3)*, 320–341.

McCann, R., & Giles, H. (2002). Ageism in the workplace: A communication perspective, 163–199.

McMullin, J. A., & Marshall, V. W. (2001). Ageism, age relations, and garment industry work in montreal. *The Gerontologist*, *41(1)*, 111–122.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58(4)*, 525–543.

Munck, I., Barber, C., & Torney-Purta, J. (2018). Measurement invariance in comparing attitudes toward immigrants among youth across europe in 1999 and 2009: The alignment method applied to IEA CIVED and ICCS. *Sociological Methods & Research*, *47*, 687–728. doi:doi:10.1177/0049124117729691

Muthén, B. O., & Asparouhov, T. (2013). BSEM measurement invariance analysis. Retrieved August 14, 2018. Retrieved from http://www.statmodel.com/

Muthén, B. O., & Asparouhov, T. (2014). IRT studies of many groups: The alignment method. *Frontiers in Psychology*, *5*, 978.

Muthén, B. O., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups. *Sociological Methods & Research*, *47*, 637–664. doi:doi:10.1177/0049124117701488

Muthén, B. O., du Toit, S. H. C., & Spisic, D. (1997). *Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes.* Unpublished manuscript.

Muthén, L., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Los Angeles, CA: Muthén & Muthén.

Palmore, E. (1999). *Ageism: Negative and positive*. New York, NY: Springer Publishing Company.

Pasupathi, M., & Löckenhoff, C. E. (2002). Ageist behavior. In T. D. Nelson (Ed.), *Ageism: Stereotyping and prejudice against older persons* (pp. 201–246). Cambridge, MA.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, *8(2)*, 287–312.

Posthuma, R. A., & Campion, M. A. (2009). Age stereotypes in the workplace: Common stereotypes, moderators, and future research directions. *Journal of Management*, *35(1)*, 158–188.

Raymer, M., Reed, M., Spiegel, M., & Purvanova, R. K. (2017). An examination of generational stereotypes

as a path towards reverse ageism. *The Psychologist-Manager Journal*, *20(3)*, 148–175.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin*, *114(3)*, 552–566.

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17(3)*, 354–373.

Rodham, K., Glover, I., & Branine, M. (2001). Ageism, young academics and the buffalo stance. In I. Glover & M. Branine (Eds.), *Ageism in work and employment*. Aldershot: Ashgate Publishing.

Rost, K., & Frankel, R. (1993). The introduction of the older patient's problems in the medical visit. *Journal of Aging and Health*, *5(3)*, 387–401.

Rupp, D. E., Vodanovich, S. J., & Credé, M. (2005). The multidimensional nature of ageism: Construct validity and group differences. *The Journal of Social Psychology*, *145(3)*, 335–362.

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. *Sociological Methodology*, *17*, 105–129.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, *7(2)*, 147–177.

Seddig, D., & Leitgöb, H. (2018a). Approximate measurement invariance and longitudinal confirmatory factor analysis: Concept and application with panel data. *Survey Research Methods*, *12(1)*, 29–41.

Seddig, D., & Leitgöb, H. (2018b). Exact and bayesian approximate measurement invariance. In E. Davidov, P. Schmidt, J. Billiet, & B. Meuleman (Eds.), *Cross-cultural analysis: Methods and applications* (2nd, pp. 553–579). New York: Routledge.

Sharps, M. J., Price-Sharps, J. L., & Hanson, J. (1998). Attitudes of young adults toward older adults: Evidence from the United States and Thailand. *Educational Gerontology*, *24(7)*, 655–660.

Singer, M. S. (1986). Age stereotypes as a function of profession. *The Journal of Social Psychology*, *126(5)*, 691–692.

Snape, E., & Redman, T. (2003). An evaluation of a three-component model of occupational commitment: Dimensionality and consequences among United Kingdom human resource management specialists. *Journal of Applied Psychology*, *88(1)*, 152–159.

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

Thompson, N. (1997). Children, death and ageism. *Child & Family Social Work*, *2(1)*, 59–65.

Tien-Hyatt, J. L. (1987). Self-perceptions of aging across cultures: Myth or reality? *The International Journal of Aging and Human Development*, *24(2)*, 129–148.

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*, 1–15.

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.

Vauclair, C.-M., Abrams, D., & Bratt, C. (2010). Measuring attitudes to age in britain: Reliability and validity of the indicators. London, UK: Department for Work and Pensions.

Walker, A. (1990). The economic "burden" of ageing and the prospect of intergenerational conflict. *Ageing and Society*, *10(4)*, 377–396.

Wearing, B. (1995). Leisure and resistance in an ageing society. *Leisure Studies*, *14*.

West, S. G., Taylor, A. B., & Wu, W. (2012). Model fit and model selection in structural equation modeling. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 209–231). New York, NY: Guilford Press.

Westman, J. C. (1991). Juvenile ageism: Unrecognized prejudice and discrimination against the young. *Child Psychiatry & Human Development*, *21(4)*, 237–256.

World Health Organization. (2015). World report on ageing and health. World Health Organization.

Xiao, L. D., Shen, J., & Paterson, J. (2013). Cross-cultural comparison of attitudes and preferences for care of the elderly among Australian and Chinese nursing students. *Journal of Transcultural Nursing*, *24(4)*, 408–416.