

Measurement Equivalence of the English, German, and French Versions of the Strike Attitude and Behavioral Reactions Scale (SABeRS)

Denise Vesper¹ and Cornelius J. König¹

Department of Psychology, Universität des Saarlandes, Saarbrücken, Germany

Abstract: Strikes are an important phenomenon in the working world. Nevertheless, cross-cultural psychological research on strikes has been limited as appropriate scales were missing. Recently, a scale to determine third-party attitudes and behavioral reactions was introduced: the Strike Attitude and Behavioral Reactions Scale (SABeRS). The applicability of this scale is currently limited to a German context, as it was developed in German. We thus decided to extend the applicability to other languages, that is, English and French. To test the measurement equivalence of the SABeRS, we used a British ($n = 444$), a German ($n = 454$), and a French sample ($n = 463$) and ran multigroup confirmatory factor analyses. Based on multi-group CFA and alignment optimization analyses, the scale was found to be at least partially measurement equivalent between the groups. The five factors were consistently confirmed in all samples. Overall, this study indicates that the SABeRS is psychometrically solid and that it is measurement-equivalent in English, German, and French samples.

Keywords: strikes, measurement equivalence, third-party, industrial relations, unions



Strikes from university lecturers, air traffic controllers, and hospital staff are three examples of strikes that could happen in many countries and that can affect the public in their everyday life. The public can be an important stakeholder in strikes, especially for unions that call for strikes. Unions build on the legitimacy and approval of the public for conducting and continuing strikes and hence, on positive attitudes of third parties to strikes (Kelloway et al., 2008). In such cases, knowing about the attitudes of the public to strikes would be helpful in the decision-making process for unions in all these countries – but also for employers, on the other side of the bargaining table.

Given the importance of public attitudes to strikes for employers and unions and thus also for everybody interested in understanding strikes and their consequences, it also becomes important to measure these attitudes. So far, only the Strike Attitude and Behavioral Reactions Scale

(SABeRS; Vesper & König, 2022) has been proposed as a German scale to measure these beliefs and behavioral reactions. The measure is based on the five factors negative reactions to strikes, legitimacy of strikes, informing oneself about strikes, behavior in social networks to strikes, and support of strikers. In their scale development paper, Vesper and König (2022) conducted four studies in Germany. Study 1 aimed at reducing the initial item pool, and Study 2 showed the reliability of the scale consisting of five factors with a new sample. In Study 3, Vesper and König (2022) assessed the convergent and discriminant validity of the scale. They found that the five factors were significantly associated with attitudes toward unions and willingness to strike and were not or to a smaller extent related to general self-efficacy, openness to new experiences, and extraversion. Negative reactions to strikes were negatively associated with readiness to strike, union attitudes/loyalty, and having a politically left view, whereas the other four factors were positively associated with these variables. Furthermore, people with no previous strike participation and non-union members indicated more negative reactions and less legitimacy of strikes, informing themselves about

strikes, strike-related social network behavior, and support of strikers compared to people who had previously participated in a strike and union members. In Study 4, Vesper and König (2022) applied the scale to a specific strike showing that the scale also works for specific and not only for general attitudes and reactions to strikes. In this study, the same pattern of correlations was found as in Study 3. Furthermore, strikers and individuals not affected by the strike reported fewer negative reactions than strike-affected third parties. Strikers also had the highest mean in all other factors compared to the other two groups.

The SABeRS could also be a useful tool for examining differences in the attitudes toward strikes within other countries as well as across countries – differences due to the different frequency, length, or sectors affected by strikes in different countries. However, a comparison of third-party attitudes and reactions toward strikes within other countries and across countries is currently not possible because there is only a German version of the SABeRS. Such a comparison between countries is only valid if the scale is measurement equivalent between the countries. Measurement equivalence ensures that the differences between countries are due to differences in the construct being measured and not due to different understanding of items by different groups of participants or improper translation (Byrne & Van de Vijver, 2010).

Research differentiates between configural, metric, and scalar measurement equivalence (Vandenberg & Lance, 2000). Configural equivalence assesses whether the examined groups used the concurring conceptual reference frame for the construct. If configural equivalence is established, the responses of each group can be divided into the same number of factors, and the same items are assigned to the respective factors (Meredith, 1993). If configural equivalence is ensured, metric equivalence can be examined. Metric equivalence exists when the data from the groups studied demonstrate similar strengths and magnitudes of the relation with the factor and hence, have the same factor loadings (Bollen, 1989). If metric equivalence is found, scalar equivalence can be tested. Scalar equivalence tests for invariance of the vector of item intercepts, with item intercepts defined as the individual items' values fitting to the zero value of the underlying construct (Meredith, 1993). The comparison of latent means requires scalar or strong factorial equivalence. Because it requires that the scales in all groups be operationally defined in the same way (Cheung & Rensvold, 2002). Thus, measurement equivalence is achieved when individuals from different groups with identical latent construct values exhibit matching expected manifest scores (Drasgow & Kanfer, 1985). Hence, we hypothesize more formally:

Hypothesis 1: The SABeRS will be measurement equivalent in the samples from the United Kingdom, Germany, and France.

Methods

Sample

Data were collected through Respondi, an online panel provider that operates online panels in seven different countries, including the United Kingdom, Germany, and France. The dataset is available at https://osf.io/46bdr/?view_only=bc86163deafd4aa880e5733c4c18e743. Participants received 0.50 € as compensation. In total, 1,652 participants completed the study. The dataset used for the analyses reported in this paper was also used for two other papers. In the first, we assessed the measurement equivalence of the General System Justification Scale (Kay & Jost, 2003; Vesper et al., 2022). The only overlapping variable between this paper and the System Justification Scale paper is political orientation. In the second paper, we relied on the results of this paper and compared the strike attitudes between the three samples. Hence, this second paper builds on those results, but reports mean comparisons (Vesper & König, 2023a). We have uploaded a data transparency table to the Open Science Framework (OSF) which further explains the similarities and differences between the three papers. Due to the collection of data on willingness to strike for the other paper, only those participants who were currently employed filled out the questionnaire; all other participants were screened out at the beginning of the survey ($n = 92$) as these can formally not conduct a work stoppage. To ensure data quality (Meade & Craig, 2012), we followed several steps. First, participants who selected “No” when asked whether we could analyze their data for scientific purposes (Meade & Craig, 2012) were excluded from analyses ($n = 33$). Second, to take care of swift completion, we excluded all participants ($n = 78$) who needed less than two seconds per item to answer the items (Huang et al., 2012). Finally, we examined long strings, which are defined as how often participants consecutively selected the same response option. Johnson (2005) recommends checking the data for a so-called “elbow.” In our data, the elbow appeared at six items, hence long strings above six items were identified ($n = 88$). The analyses below were conducted without the data from participants with long strings (Johnson, 2005; Niessen et al., 2016). With this exclusion procedure, we followed the information provided in the preregistration for this study (<https://aspredicted.org/xf74z.pdf>).¹ After controlling for these aspects, $N = 1,361$ people were included in the analyses.

¹ The preregistration includes additional hypotheses that go beyond the measurement equivalence analyses that are the focus of this paper.

Table 1. Correlations between the SABeRS, union membership, strike participation, and political orientation

Variables	Negative reactions toward strikes	Legitimacy of strikes	Informing oneself about strikes	Strike-related social network behavior	Support of strikers
	UK/DE/FR	UK/DE/FR	UK/DE/FR	UK/DE/FR	UK/DE/FR
Union membership	.22/.25/.29	-.39/-.25/-.29	-.16/-.20/-.28	-.18/-.06/-.33	-.36/-.29/-.30
Previous strike participation	.44/.28/.43	-.39/-.36/-.48	-.25/-.36/-.47	-.27/-.31/-.24	-.43/-.43/-.52
Political orientation	.29/.07/.33	-.38/-.09/-.32	-.24/-.13/-.26	-.17/-.05/-.12	-.33/-.16/-.36

Note. $N_{UK} = 436$, $N_{DE} = 447$, $N_{FR} = 454$. UK = United Kingdom; DE = Germany; FR = France. Union membership was coded with 0 = union member, 1 = no union member. Previous strike participation was coded with 0 = previous participation, 1 = no previous participation. Political orientation was coded from 1 = left to 11 = right. Values in bold are significant with $p < .001$. Biserial correlations are reported for the relations between union membership, strike participation, and the five factors of the SABeRS.

For the overall sample, participants were on average 46.33 ($SD = 10.03$) years old, 66.9% were female, and 33.1% were male. In total, 17.4% were union members, and 28.7% had already participated in a strike. For the UK sample ($n = 444$), participants were on average 46.82 ($SD = 10.68$) years old. Of the British participants, 65.8% were female and 34.2% were male. In the United Kingdom sample, 22.5% were union members, and 18.2% had previously participated in a strike. For the sample from Germany ($n = 454$), participants were on average 44.80 ($SD = 10.64$) years old, 65.4% were female, and 34.6% were male. In the German sample, 13.7% were union members, and 21.8% had participated in a strike themselves. French participants ($n = 463$) were on average 47.36 ($SD = 8.53$) years old. More than two-third (69.3%) reported being female and 30.7% reported being male. In the French sample, 16.2% were union members, and 45.4% had participated in a strike.

Translation Process

We translated the German SABeRS and the items on willingness to strike (Vesper & König, 2022) into English and French using a back-translation process based on recommendations from Schaffer and Riordan (2003). For both English and French, we consulted two individuals, one of these was either a native or fluent speaker in English or French, respectively, as well as German, whereas the second individual was a native German speaker and trained translator in one of the respective languages. Differences that occurred between translated versions were discussed between the translators, who then decided on the appropriate translation. The German, English, and French items are uploaded to the OSF.

Materials

We used the 15-item SABeRS with three items per factor. Items were rated on a 5-point Likert scale ranging from 1 = *do not agree* to 5 = *agree*.

Statistical Analyses

Statistical analyses were conducted using R 3.6.1 (R Core Team, 2019) and several R packages: *careless* (Yentes &

Wilhelm, 2018), *dmacs* (Dueber, 2019), *lavaan* (Rosseel, 2012), *MBESS* (Kelley, 2022), *sem* (Fox et al., 2017), *sem-Tools* (Jorgensen et al., 2019), and *sirt* (Robitzsch, 2021). The comparative fit index (CFI), the Tucker-Lewis index (TLI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR) were used to assess fit in confirmatory factor analyses (CFAs). To assess model fit, we relied on the following recommendations of Hu and Bentler (1999) who see CFI and $TLI \geq .95$, $SRMR \leq .08$, and $RMSEA \leq .06$ as indicating values for good model fit. Correlations between the five factors of the SABeRS, union membership, previous strike participation, and political orientation can be found in Table 1.

For assessing measurement equivalence, we followed the simultaneous approach using multigroup CFAs (Somaraju et al., 2022). We decided to compare the German sample with the British and the French, respectively, because the original language of the scale was German. Before the first step of this simultaneous approach, two preliminary analyses had to be done: separate CFAs for each sample and the definition of a baseline model for the multigroup CFA, with the latter consisting of similar loading patterns for all groups, whereas the magnitude of loadings, intercepts, variances, factor covariances, construct means, and residual variances are allowed to vary. This free baseline model (Stark et al., 2006) is then used to assess the configural equivalence as the first step of the multigroup CFA. Hence, to evaluate the configural equivalence the same number of latent variables and the same loading patterns of the latent variables on the indicators across the examined groups are specified. Second, metric equivalence, which aims at ensuring similar magnitudes of factor loadings and regression weights (from the factors to items) across groups, and scalar equivalence, the invariance of the vector of the item intercepts, are tested simultaneously. If scalar equivalences are established, the latent variables' means can be compared meaningfully across the examined groups (Chen, 2008). Changes in CFI of .002 (or less) when comparing a model to a less constrained model indicate that the equivalence hypothesis should not be rejected (Meade et al., 2008).

Results

Preliminary Analyses

We first conducted separate CFAs for each sample to test the fit of the proposed five-factor model before conducting the multigroup CFA, as suggested by Sass (2011). The CFA for the sample from the United Kingdom showed no good model fit of the five-factor model based on the recommendations of Hu and Bentler (1999), $\chi^2(80) = 306.75$, $p < .001$, CFI = .94, TLI = .92, RMSEA = .08, 90% CI [.07, .09], SRMR = .08. However, the CFI, TLI, and RMSEA values were still close to their respective cut-offs and the SRMR even lay within the cut-off. Furthermore, the cut-off criteria from Hu and Bentler (1999) are not without criticism as they were overgeneralized in their use (Hu & Bentler, 1998; McNeish, 2023). Methodological studies found that cut-off values can also change depending on model characteristics and data (see McNeish & Wolf, 2023, p. 62 for an overview). Hence, the model can be assumed to be at least satisfactorily fitting the data. The five-factor model fitted the data significantly better than a one-factor model, $\Delta\chi^2(10) = 617.76$, $p < .001$.

For the German sample, the fit of the five-factor model was good, with only the TLI and RMSEA lying slightly below the cut-off criteria from Hu and Bentler (1999), $\chi^2(80) = 254.12$, $p < .001$, CFI = .95, TLI = .94, RMSEA = .07, 90% CI [.06, .08], SRMR = .05. The assumed five-factor model fitted the data significantly better than a one-factor model, $\Delta\chi^2(10) = 863.84$, $p < .001$. Finally, the five-factor model showed a good model fit in the French sample as well. In this model, CFI and SRMR met the cut-off criteria, whereas TLI and RMSEA slightly exceeded the cut-off criteria, $\chi^2(80) = 293.23$, $p < .001$, CFI = .95, TLI = .94, RMSEA = .08, 90% CI [.07, .09], SRMR = .06. In this sample the five-factor model fitted the data again significantly better than a one-factor model, $\Delta\chi^2(10) = 641.74$, $p < .001$. Figure 1 gives an overview of the three CFAs.

Before specifying the baseline model, we chose the referent item for each factor and each comparison. For this, we assessed which item exhibited the highest equivalence following recommendations from Nye and Drasgow (2011). During the analyses, the referent items' factor loading was set to 1 and the intercept to 0 (Somaraju et al., 2022).

Furthermore, we specified a baseline model for each comparison. The baseline model was calculated combining the two respective samples and was otherwise similar to the later used configural model. The fit indices of this baseline model for the German-British comparison were close to the cut-off criteria from Hu and Bentler (1999), indicating no good, but at least satisfactory fit, $\chi^2(80) = 490.43$, $p < .001$, CFI = .941, TLI = .92, RMSEA = .08, 90% CI [.07, .08], SRMR = .06. For the German-French comparison,

the fit indices also exhibited a satisfactory fit, $\chi^2(80) = 492.99$, $p < .001$, CFI = .94, TLI = .93, RMSEA = .08, 90% CI [.07, .08], SRMR = .05. This shows that the groups display similar loading patterns, while the magnitude of loadings, intercepts, variances, factor covariances, construct means, and residual variances are allowed to vary.

Test of Hypothesis

To test for measurement equivalence of the SABeRS between the German and British sample and the German and French sample, we followed two analytic steps (see also above). As a first step, we tested configural equivalence, and results showed only satisfactory model fit with all indices not meeting the cut-off criteria in the German-British comparison, $\chi^2(160) = 560.87$, $p < .001$, CFI = .944, TLI = .93, RMSEA = .08, 90% CI [.07, .08], SRMR = .06, and the German-French comparison, $\chi^2(160) = 547.35$, $p < .001$, CFI = .952, TLI = .94, RMSEA = .07, 90% CI [.07, .08], SRMR = .05. As the CFI, TLI, and RMSEA were close to the required cut-offs, we evaluated configural equivalence as given.

In the second step, we assessed the metric and scalar equivalence simultaneously. The model fit indices were all close to the cut-off criteria, indicating no good, but at least satisfactory fit for the German-British comparison, $\chi^2(180) = 667.72$, $p < .001$, CFI = .932, TLI = .92, RMSEA = .08, 90% CI [.07, .08], SRMR = .07, and the German-French comparison, $\chi^2(180) = 771.82$, $p < .001$, CFI = .926, TLI = .91, RMSEA = .09, 90% CI [.08, .09], SRMR = .07. Compared with the configural model, the change in CFI was $\Delta\text{CFI} = -.012$ for the German-British comparison and $\Delta\text{CFI} = -.025$ for the German-French comparison. The scale was hence neither fully scalar measurement equivalent in the German-British comparison nor in the German-French comparison based on the threshold of $\Delta\text{CFI} = .002$ (Meade et al., 2008).

To further examine scalar equivalence, we decided to take a closer look at which parameters should be released and to test for partial scalar equivalence by removing the constraints on item intercepts based on modification indices and d_{MACS} and retesting the model (Putnick & Bornstein, 2016). For the British-German comparison, the items "Strikes strain myself," "Strikes are a waste of time," and "I would support the strikers' position in conversations" had the highest modification indices and the highest d_{MACS} (0.47, 0.38, and 0.27, respectively). Releasing these items from their constraints leads to an improvement in model fit. The ΔCFI was however still above the recommended cut-off with $\Delta\text{CFI} = -.005$. We therefore further released the constraints from the items "I am interested in the reasons for strikes" and "I take a look at posts about

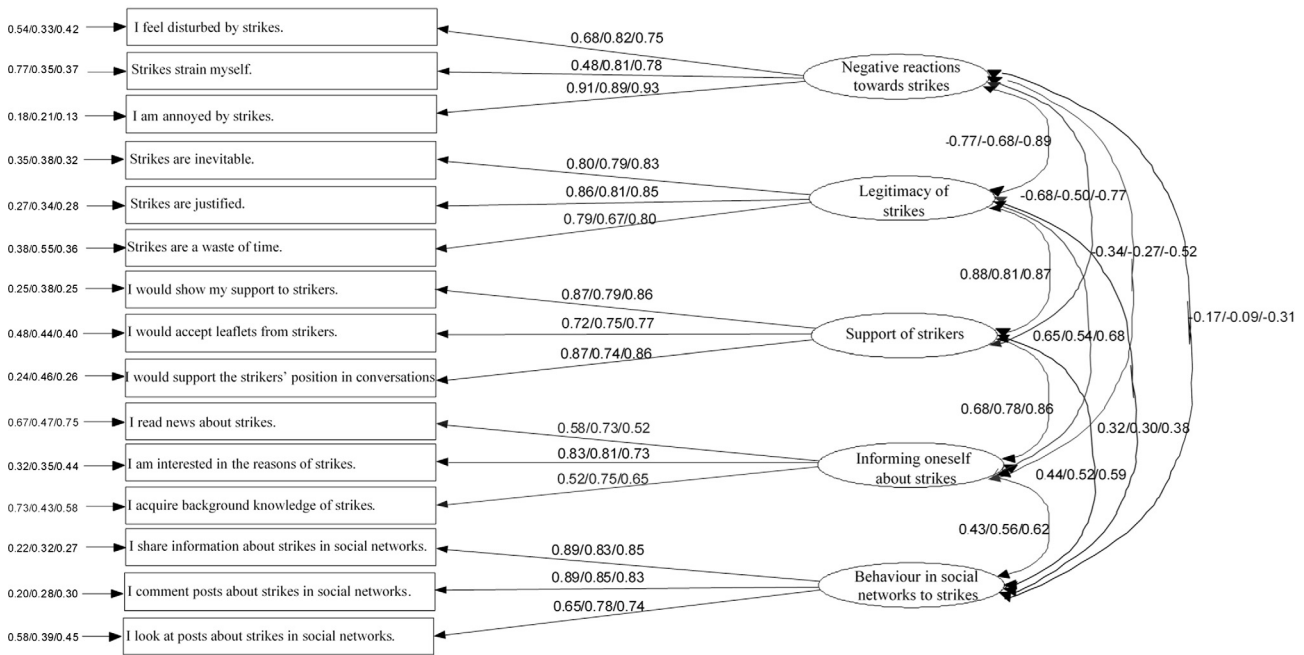


Figure 1. Results of the confirmatory factor analyses of the three samples. Numbers represent standardized loadings. The order of the results is British/German/French.

strikes on social networks.” This improved the model fit further. The cut-off of $\Delta CFI = .002$ was however still not met with $\Delta CFI = -.003$. In a final step, we then released the constraints from the item “Strikes are necessary.” This changed the ΔCFI to $\Delta CFI = .001$ and hence below the cut-off. However, for the legitimacy of strikes factor two of three parameters were now released from their constraints. Hence, there is only partial support for the assumption of measurement equivalence between the German and English scales.

Regarding the German-French comparison, the items “Strikes are a waste of time” and “I would show my support to strikers” showed the highest modification indices and the highest d_{MACS} (-0.57 and 0.38 , respectively). Therefore, we loosened the constraints on these items and tested for scalar equivalence with the adapted model. The adapted model had a significantly better fit compared to the scalar equivalence model ($\Delta CFI = -.006$). It was however still above the recommended threshold of $\Delta CFI = .002$. Therefore, we released the items constraints of the items “I read news about strikes” and “I share information about strikes on social media” after consulting the modification indices. This led to $\Delta CFI = .002$. Hence, partial scalar equivalence could be achieved. Following recommendations from Steenkamp and Baumgartner (1998) and Vandenberg and Lance (2000) we assumed that the scale was partially equivalent as the majority of items on the factors were equivalent. Hence, partial scalar equivalence for the five-factor model

was obtained in the German-French comparison, offering partial support for our hypothesis.

The multigroup factor analyses were conducted as pre-registered. However, we added the alignment optimization analyses to support our results, which were not mentioned in the pre-registration. To back up our results, we also performed the measurement equivalence test using the alignment optimization method (Asparouhov & Muthén, 2014; Magraw-Mickelson et al., 2020), which assumes approximate rather than exact invariance (see Magraw-Mickelson et al., 2020 for a detailed description of this method). Therefore, this method allows for some degree of non-invariance between the groups. Less than 20% of non-invariant parameters are considered to be acceptable (Asparouhov & Muthén, 2014). The degree of non-equivalence between the parameters was 0% for three factors. Only the factors of the legitimacy of strikes and support of strikes had non-equivalent item parameters. However, only 11.1% of item parameters were non-equivalent for both factors, which is below the recommended threshold.

We calculated d_{MACS} as a non-equivalence effect size for the scales (Nye & Drasgow, 2011). These d_{MACS} are defined as the proportion that differential item functioning (DIF) has on the expected score differences for each item. The effect sizes are defined as follows: 0.40 is a small, 0.60 is a medium, and 0.80 is a large effect (Nye et al., 2018). In our case, we chose the German sample as the referent group as the scale was originally developed in German. On the item level,

Table 2. Effect sizes (d_{MACS}) of the measurement non-equivalence of the SABeRS

Item	d_{MACS}	Item $\Delta mean$	d_{MACS}	Item $\Delta mean$
	UK ^a	UK	France ^a	France
I feel disturbed by strikes.	–	–	0.20	–0.20
Strikes strain myself.	0.47	–0.08	–	–
I am annoyed by strikes.	0.30	–0.11	0.22	–0.19
Strikes are necessary.	0.31	0.25	–	–
Strikes are justified.	–	–	0.24	–0.04
Strikes are a waste of time. (reverse-coded)	0.38	0.25	0.57	0.47
I read news about strikes.	0.22	–0.17	0.21	–0.18
I am interested in the reasons for strikes.	0.21	–0.09	–	–
I acquire background knowledge about strikes myself.	–	–	0.08	0.07
I share information about strikes on the social media.	–	–	0.24	–0.15
I comment on posts about strikes on the social media.	0.04	–0.04	0.12	–0.09
I take a look at posts about strikes on social networks.	0.16	–0.08	–	–
I would show my support to strikers.	–	–	0.38	0.35
I would accept flyers from strikers.	0.07	–0.07	0.08	–0.08
I would support the strikers' position in conversations.	0.27	–0.19	–	–

Note. d_{MACS} (with MACS = mean and covariance structure) = effect size for measurement non-equivalence on the item level. Items without values were referent items. Numbers in bold represent small effects ($> .40$). Reference group was Germany.

Table 3. Effects of non-equivalence on scale-level properties

Scale	$\Delta mean$	Observed mean difference	% of observed mean difference	d_{MACS}
Negative reactions to strikes				
DE ^a -UK	–0.19	–0.79	24%	0.47 , 0.30
DE ^a -FR	–0.39	–1.03	38%	0.20, 0.22
Legitimacy of strikes				
DE ^a -UK	0.51	0.90	57%	0.31, 0.38
DE ^a -FR	0.44	0.41	107%	0.24, 0.57
Informing oneself about strikes				
DE ^a -UK	–0.27	–0.24	113%	0.22, 0.21
DE ^a -FR	–0.11	–0.15	73%	0.20, 0.08
Strike-related social network behavior				
DE ^a -UK	–0.12	–0.23	52%	0.04, 0.16
DE ^a -FR	–0.24	–0.32	75%	0.24, 0.12
Support of strikers				
DE ^a -UK	–0.26	–0.46	57%	0.07, 0.27
DE ^a -FR	0.27	0.14	193%	0.38, 0.08

Note. d_{MACS} (with MACS = mean and covariance structure) = effect size for measurement non-equivalence on the item level. Negative values in $\Delta mean$ indicate that DIF results in higher means for the focal group compared to the referent group. UK = United Kingdom; DE = Germany; FR = France. Numbers in bold represent small effects ($> .40$). ^aReference group.

the magnitude of effects of non-equivalence ranged from 0.04 (“I comment on posts about strikes on the social media” in the German-British comparison) to 0.57 (“Strikes are a waste of time” in the German-French comparison; Table 2). The quantity of observed difference that was attributable to DIF on the scale level ($\Delta mean$) ranged from $d = -0.11$ for the comparison of informing oneself about strikes in the German and French samples to $d = 0.51$ for the comparison of the legitimacy of the strike factor among the British and German samples (Table 3). The percentage of the

observed mean difference that was traced back to DIF ranged from 24% (factor negative reactions to strikes in the British-German comparison) to 193% (factor support of strikers in the German-French comparison), with percentages larger than 100% indicating that the DIF effects are greater than the observed mean differences. This was the case in three comparisons. Overall, the effects of non-equivalence were rather small, indicating that the scale can be used for further comparisons. Nevertheless, the respective items should be considered for further improvements.

Discussion

This study assessed the measurement equivalence of the SABeRS (Vesper & König, 2022) for a British, a German, and a French sample. We followed recommendations for measurement equivalence assessment proposed by Somaraju and colleagues (2022) and showed that the SABeRS was partially scalar equivalent in the German–British comparison and in the German–French comparison. The factor structure of the SABeRS was validated in all three samples. Furthermore, the model of the five-factor structure showed satisfactory to good fit according to the cut-off criteria specified by Hu and Bentler (1999) in all three samples and the effect sizes of non-equivalence were of rather small sizes. We also obtained similar results to Vesper and König (2022) regarding the correlations of the five factors with union membership, previous strike participation, and political orientation in all three samples (see Table 1). We can thus conclude that the scale seems to be sufficiently measurement-equivalent once some adjustments have been made to the English- and French-language versions.

The SABeRS consists of the five factors negative reactions toward strikes, legitimacy of strikes, informing oneself about strikes, strike-related social network behavior, and support of strikers. Differentiating between these can help improve our perception of the concept of strike attitudes and behavioral reactions: As an example, people may report negative reactions to strikes but at the same time support strikers. In addition, people may have negative reactions to strikes but still value strikes as legitimate. We also found significant correlations indicating that union members, individuals with previous strike participation, and politically left-oriented individuals reported fewer negative reactions and more legitimacy, informing themselves more, and more support in all three samples.

The fact that the scale was only partially scalar equivalent in the British–German comparison and German–French comparison indicates that some of the items might function differently in these samples. One reason for this could be the different strike contexts of the countries. In France, strikes are more common and an individual right (Guedes & Balanescu, 2021). Whereas strikes in Germany are more regulated, for example, in that only unions are allowed to call for strikes (Büttgen & Clauwaert, 2021). Hence, these differences could also lead to different experiences with strikes and thus also to differential item function of the SABeRS. Researchers using the scale should thus always consider the context of their data collection.

Although the SABeRS showed only partially scalar equivalence in both comparisons using the multigroup CFA approach, which was achieved by loosening the constraints for several items, the alignment optimization method only flagged two items as noninvariant: “Strikes are a waste of

time” and “I would show my support to strikers.” A possible option would be to reformulate these two items. For the reverse coded item “Strikes are a waste of time” an alternative could be “The time for strikes is used wisely.” Moreover, partial scalar measurement equivalence could also be sufficient for the further use of the scale: Byrne et al. (1989) argued that latent means can be compared under partial intercept or scalar equivalence as the non-equivalent item should not affect the latent means comparison to a great extent. Moreover, Schmitt and Ali (2014) argue that next to the statistical implications, one should also consider the practical implications of research findings. They show that even in instances in which there was a relatively great lack of measurement equivalence, the practical impact of these lacks was minor. This is also supported by Leitgöb et al. (2023), who state that full invariance is almost always violated at least to some extent. Hence, as full configural equivalence, full metric equivalence, and partial scalar invariance have been established, we assume that the scale is sufficiently measurement-equivalent in the tested language versions. This can also be seen regarding the similar correlation patterns with union membership, strike participation, and political orientation in the three samples.

Limitations

The main limitation of this study is that the samples might not be fully representative of their respective country, as they were acquired via an online-panel provider and participants received a small monetary reward for their participation. Moreover, as these participants are used to filling out different questionnaires, they might have not filled out the survey as conscientiously as possible. However, by taking care of swift completion with the use of long strings and seconds used per item rates, we tried to minimize at least this possibility. Nevertheless, it could be useful to validate the scale further in a representative sample for each country. This also aligns with the limitation that we only included employed participants in our sample. To assess the strike attitudes of the public, future research should also include participants who are currently not employed, retired, or have another employment status.

Conclusion

The objective of this study was to examine the measurement equivalence of the SABeRS in English, German, and French. Our results indicate that the five factors of negative reactions toward strikes, legitimacy of strikes, informing oneself about strikes, strike-related social network behavior, and support of strikers can be found consistently in all three

language versions. The support of the (partial) measurement equivalence in these three language versions can be seen as an important step to allow for psychology-inspired cross-cultural strike research and thus to enrich the literature on strikes.

References

- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495–508. <https://doi.org/10.1080/10705511.2014.919210>
- Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley.
- Büttgen, N., & Clauwaert, S. (2021). *Germany – The right to strike in the public services (Right to strike country factsheet)*. EPSU and ETUI. https://www.epsu.org/sites/default/files/article/files/Germany%20-%20Right%20to%20strike%20in%20the%20public%20sector%20-%20EPSU%20format%20-%20Updated%20DB_0.pdf
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, 105(3), 456–466. <https://doi.org/10.1037/0033-2909.105.3.456>
- Byrne, B. M., & Van de Vijver, F. J. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, 10(2), 107–132. <https://doi.org/10.1080/15305051003637306>
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, 95(5), 1005–1018. <https://doi.org/10.1037/a0013193>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Drasgow, F., & Kanfer, R. (1985). Equivalence of psychological measurement in heterogeneous populations. *Journal of Applied Psychology*, 70(4), 662–680. <https://doi.org/10.1037/0021-9010.70.4.662>
- Dueber, D. (2019). *dmacs: Measurement nonequivalence effect size calculator (0.1.0)* [Computer software]. <https://github.com/ddueber/dmacs>
- Fox, J., Zhenghua, N., & Byrnes, J. (2017). *sem: Structural equation models (3.1-9)* [R package version 3.1-9]. <https://CRAN.R-project.org/package=sem>
- Guedes, C., & Balanescu, D. (2021). *France – The right to strike in the public services (Right to strike country factsheet)*. EPSU and ETUI. <https://www.epsu.org/sites/default/files/article/files/France%20-%20Right%20to%20strike%20in%20the%20public%20sector%20-%20updated%202021.pdf>
- Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3(4), 424–452. <https://doi.org/10.1037/1082-989X.3.4.424>
- 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. <https://doi.org/10.1080/10705519909540118>
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., & DeShon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *Journal of Business and Psychology*, 27(1), 99–114. <https://doi.org/10.1007/s10869-011-9231-8>
- Johnson, J. A. (2005). Ascertaining the validity of individual protocols from web-based personality inventories. *Journal of Research in Personality*, 39(1), 103–129. <https://doi.org/10.1016/j.jrp.2004.09.009>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2019). *semTools: Useful tools for structural equation modeling* (R package version 0.5–2) [Computer software]. <https://CRAN.R-project.org/package=semTools>
- Kay, A. C., & Jost, J. (2003). Complementary justice: Effects of “poor but happy” and “poor but honest” stereotype exemplars on system justification and implicit activation of the justice motive. *Journal of Personality and Social Psychology*, 85(5), 823–837. <https://doi.org/10.1037/0022-3514.85.5.823>
- Kelley, K. (2022). *MBESS: The MBESS R Package (4.9.1)* [Computer software]. <https://CRAN.R-project.org/package=MBESS>
- Kelloway, E. K., Francis, L., Catano, V. M., & Dupré, K. E. (2008). Third-party support for strike action. *Journal of Applied Psychology*, 93(4), 806–817. <https://doi.org/10.1037/0021-9010.93.4.806>
- Leitgöb, H., Seddig, D., Asparouhov, T., Behr, D., Davidov, E., De Roover, K., Jak, S., Meitinger, K., Menold, N., Muthén, B., Rudnev, M., Schmidt, P., & van de Schoot, R. (2023). Measurement invariance in the social sciences: Historical development, methodological challenges, state of the art, and future perspectives. *Social Science Research*, 110, Article 102805. <https://doi.org/10.1016/j.ssresearch.2022.102805>
- Magraw-Mickelson, Z., Carrillo, A. H., Weerabangsa, M. M., Owuamalam, C., & Gollwitzer, M. (2020). *Comparing classic and novel approaches to measurement invariance*. <https://doi.org/10.31234/osf.io/pz8u9>
- McNeish, D. (2023). Generalizability of dynamic fit index, equivalence testing, and hu & bentler cutoffs for evaluating fit in factor analysis. *Multivariate Behavioral Research*, 58(1), 195–219. <https://doi.org/10.1080/00273171.2022.2163477>
- McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, 28(1), 61–88. <https://doi.org/10.1037/met0000425.supp>
- Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods*, 17(3), 437–455. <https://doi.org/10.1037/a0028085>
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93(3), 568–592. <https://doi.org/10.1037/0021-9010.93.3.568>
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525–543. <https://doi.org/10.1007/BF02294825>
- Niessen, A. S. M., Meijer, R. R., & Tendeiro, J. N. (2016). Detecting careless respondents in web-based questionnaires: Which method to use? *Journal of Research in Personality*, 63, 1–11. <https://doi.org/10.1016/j.jrp.2016.04.010>
- Nye, C. D., Bradburn, J., Olenick, J., Bialko, C., & Drasgow, F. (2018). How big are my effects? Examining the magnitude of effect sizes in studies of measurement equivalence. *Organizational Research Methods*, 22(3), 678–709. <https://doi.org/10.1177/1094428118761122>
- Nye, C. D., & Drasgow, F. (2011). Effect size indices for analyses of measurement equivalence: Understanding the practical importance of differences between groups. *Journal of Applied Psychology*, 96(5), 966–980. <https://doi.org/10.1037/a0022955>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- R Core Team. (2019). *R: A language and environment for statistical computing (3.6.1)* [Computer software]. R Foundation for Statistical Computing. <https://www.R-project.org/>

- Robitzsch, A. (2021). *sirt: Supplementary item response theory models* (R package version 3.11-21) [Computer software]. <https://CRAN.R-project.org/package=sirt>
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5-12 (BETA). *Journal of Statistical Software*, 48, 1–36.
- Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment*, 29(4), 347–363. <https://doi.org/10.1177/0734282911406661>
- Schaffer, B. S., & Riordan, C. M. (2003). A review of cross-cultural methodologies for organizational research: A best-practices approach. *Organizational Research Methods*, 6(2), 169–215. <https://doi.org/10.1177/1094428103251542>
- Schmitt, N., & Ali, A. A. (2014). The practical importance of measurement invariance. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 327–346). Routledge.
- Somaraju, A. V., Nye, C. D., & Olenick, J. (2022). A review of measurement equivalence in organizational research: What's old, what's new, what's next? *Organizational Research Methods*, 25(4), 741–785. <https://doi.org/10.1177/10944281211056524>
- Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology*, 91, 1292–1306. <https://doi.org/10.1037/0021-9010.91.6.1292>
- Steenkamp, J.-B. E., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25(1), 78–90. <https://doi.org/10.1086/209528>
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4–70. <https://doi.org/10.1177/109442810031002>
- Vesper, D., & König, C. J. (2022). Ever thought about strikes? Development of a scale to assess attitudes and behavioral reactions to strikes. *Journal of Business and Psychology*, 37(6), 1275–1298. <https://doi.org/10.1007/s10869-022-09801-7>
- Vesper, D., & König, C. J. (2023a). Differences in strike attitudes and behavioural reactions among British, German, and French samples. *Workers of the World*, 1(11), 58–91.
- Vesper, D., & König, C. J. (2023b, December 5). *Measurement equivalence of the SABeRS* [Data]. <https://osf.io/46bdr>
- Vesper, D., König, C. J., Siegel, R., & Friese, M. (2022). Is use of the general system justification scale across countries justified? Testing its measurement equivalence. *British Journal of Social Psychology*, 61(3), 1032–1049. <https://doi.org/10.1111/bjso.12520>
- Yentes, R. D., & Wilhelm, F. (2018). *careless: Procedures for computing indices of careless responding* (R package version 1.1.3) [Computer software]. <https://github.com/ryentes/careless/>

History

Received August 4, 2022

Revision received July 28, 2023

Accepted September 14, 2023

Published online December 31, 2023

EJPA Section I/O Psychology

Conflict of Interest

The authors declare no conflict of interest.

Publication Ethics

Ethical approval was not required for these studies in accordance with the local legislation and institutional requirements of Saarland University.

Open Science

We report all data exclusions, all data exclusion criteria (as established prior to data analysis in the preregistration at <https://aspredicted.org/xf74z.pdf>), and all analyses, including all tested models.

Open Data: The dataset is available at <https://osf.io/46bdr/> (Vesper & König, 2023b). We confirm that there is sufficient information for an independent researcher to reproduce all reported results, including a codebook and the code.

Open Materials: We confirm that there is sufficient information for an independent researcher to reproduce all the reported methodology.

Preregistration and Analysis Plans: This study was preregistered at <https://aspredicted.org/xf74z.pdf>.

Funding

The authors did not receive funding for this research. Open access publication was enabled by Universität des Saarlandes, Germany.

ORCID

Denise Vesper

 <https://orcid.org/0000-0002-1585-9243>

Cornelius J. König

 <https://orcid.org/0000-0003-0477-8293>

Denise Vesper

Department of Psychology

Universität des Saarlandes

Campus A1 3

66123 Saarbrücken

Germany

denise.vesper@uni-saarland.de