Work, Aging and Retirement, 2025, **XX**, XX–20
https://doi.org/10.1093/workar/waaf006
Empirical Article

OXFORD

Empirical Article

Development and validation of the short form of the Later Life Workplace Index: a study across 10 countries

Julia S. Finsel[1], Hila Axelrad[2], Su Jung Choi[3,4], Eva Derous[5,6], Xiuzhu Gu[7], Paolo Guandalini[8], Jaeyoung Ha[4], Eun Seok Kim[9], Izabela Marzec[10], Reidar J. Mykletun[11], Eduardo Oliveira[12], Sofija Pajic[13], Maaike Schellaert[5], Beatrice I.J.M. Van der Heijden[13,14,15,16,17], Michela Vignoli[8], Anne M. Wöhrmann[1,18], and Jürgen Deller[1]

[1]Institute for Management and Organization, Leuphana University of Lüneburg, Lüneburg, Germany
[2]Aaron Institute for Economic Policy, Tiomkin School of Economics, Reichman University, Herzliya, Israel
[3]Department of Vocational Education and Workforce Development, Seoul National University, Seoul, Republic of Korea
[4]Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul, Republic of Korea
[5]Vocational and Personnel Psychology Lab, Department of Work, Organization and Society, Ghent University, Ghent, Belgium
[6]Work and Organizational Psychology, Erasmus School of Social and Behavioural Science, Erasmus University Rotterdam, Rotterdam, The Netherlands
[7]Department of Industrial Engineering and Economics, School of Engineering, Institute of Science Tokyo, Tokyo, Japan
[8]Department of Psychology and Cognitive Science, University of Trento, Rovereto, Italy
[9]Korea Employment Information Service, Chungcheongbuk-do, Republic of Korea
[10]Department of Digital Economy Research, University of Economics in Katowice, Katowice, Poland
[11]UiS School of Business and Law, University of Stavanger, Stavanger, Norway
[12]School of Economics and Management, Center for Economics and Finance at UPorto, University of Porto, Porto, Portugal
[13]Institute for Management Research, Radboud University, Nijmegen, The Netherlands
[14]Faculty of Management, Open Universiteit, Heerlen, The Netherlands
[15]Department of Marketing, Innovation and Organisation, Ghent University, Ghent, Belgium
[16]The University of the Free State Business School, The University of the Free State, Bloemfontein, South Africa
[17]Kingston Business School, Kingston University, London, United Kingdom
[18]Federal Institute for Occupational Safety and Health (BAuA), Berlin, Germany

Corresponding author: Institute for Management and Organization, Leuphana University of Lüneburg, Universitätsallee 1, 21335 Lüneburg, Germany. Email: julia.finsel@leuphana.de

Abstract

This study aims to develop and validate a short form of the Later Life Workplace Index (LLWI). The LLWI measures organizational practices relevant to older workers' ability, motivation, and opportunity to continue working in later life. It consists of nine domains comprising three to four conceptual indicators each. We applied a combined qualitative and quantitative strategy to reduce the original LLWI 80-item measure into a 29-item short form (LLWI-SF). Different language versions were developed and applied to samples in ten countries: Belgium ($N = 444$), Germany ($N = 387$), Italy ($N = 408$), Japan ($N = 349$), Korea ($N = 350$), Norway ($N = 140$), Poland ($N = 353$), Portugal ($N = 306$), the Netherlands ($N = 317$), and the United States ($N = 370$). Our data provides evidence for the reliability and validity of the LLWI-SF in the ten countries and languages. Moreover, partial measurement invariance of the LLWI-SF is supported despite the diverging regulatory and cultural contexts. Thus, we provide a short but holistic measure of organizational practices for older workers that can be efficiently used in research and practice.

Keywords: human resource management, scale development/measurement/validation, cross-cultural issues, leadership

In light of demographic developments, it becomes increasingly crucial for societies and organizations worldwide to prolong labor force participation while protecting the career sustainability of individual workers (de Vos et al., 2020; Hertel & Zacher, 2018; Pilipiec et al., 2021; Riekhoff, 2024). An increasing share of older persons are deciding to continue working until or even beyond the normal retirement age (OECD, 2023). In order to protect and further enhance the career sustainability of workers into their later life and to prevent early retirement, organizational practices able to respond to the abilities, motivations, needs, and values of older workers are essential (Boehm et al., 2021; Farr-Wharton et al., 2023; Hallpike et al., 2025; Jonsson et al., 2020). Consequently, there has been an increasing interest in studying organizational practices and their implications for older workers (Boehm et al., 2021; Marcus et al., 2024).

Psychometrically sound (i.e., reliable and valid) measures are needed as a starting point for purposeful research

Received March 5, 2024; revisions received April 3, 2025; accepted April 15, 2025

to shed light on the prevalence and predictive validity of such organizational practices and enable researchers to develop evidence-based recommendations for practitioners. As demonstrated by Boehm et al. (2021) in their review of the literature, organizational practices for older workers constitute a multidimensional construct that includes age-specific, age-inclusive, and age-neutral practices. As such, the construct covers a multitude of different domains of organizational practices, starting from recruitment and selection (e.g., age-neutral language in job advertisements) to retirement and bridge employment (e.g., flexible transition to retirement and continued employment in retirement). This breadth, however, is seldomly reflected in the available measures. Existing measures tend to capture specific areas of organizational practices (e.g., age-diversity climate scale; Boehm et al., 2014) or use a noncomprehensive unidimensional scale (e.g., available HR practices scale; Taneva & Arnold, 2018). A comprehensive measure is needed to allow for a better assessment of the organizational situation, enabling better predictions of individual and organizational outcomes.

The 80-item multidimensional Later Life Workplace Index (LLWI) captures these organizational practices for older workers in a comprehensive way (Wilckens et al., 2021). It comprises nine domains: age-friendly organizational climate, leadership, work design, health management, individual development, knowledge management, transition to retirement, continued employment in retirement, and health and retirement coverage (see Figure 1 for an overview).

Using the LLWI to measure organizational practices for older workers yields several benefits for researchers and practitioners. First, the LLWI constitutes a unique holistic measure that taps into the full employee life cycle from recruitment to retirement as well as all areas of organizational practices that address older workers' ability, motivation, needs, and values, all of which influence their likelihood to continue working in later life (Wilckens et al., 2021). Second, the LLWI is flexible in its application (Finsel et al., 2023c; Wilckens et al., 2021). The measure can be used as a comprehensive tool, or domain-specific scales can be selected to focus on specific topics. Since its items focus on the organizational level, it can also be used to assess perceptions of these organizational practices by any organizational stakeholder (e.g., employees, line managers, or human resource [HR] managers). Third, the LLWI has been developed independently of the requirements of any particular industry, making it applicable to all sectors of the economy. We recommend applying this measure in organizations with at least 30 employees to ensure a meaningful assessment, giving the organization the opportunity to be actively and systematically involved in HR management (Wilckens et al., 2021).

Even though the LLWI is a fairly new measure, it has already been applied in emerging research across several countries and occupational contexts. In their quantitative studies, researchers from Bulgaria (Stoyanova et al., 2024), Germany (Finsel et al., 2023a; Wilckens et al., 2023), and Israel (Axelrad et al., 2024) have selected subscales of the LLWI to assess specific domains as predictors of older workers' outcomes or age stereotypes. For their qualitative study, researchers from Switzerland (Froidevaux et al., 2024) have also used selected domains. Researchers have further used some of the LLWI items as an inspiration in the development of an Autonomy-Support HR Practices Scale (Laguerre & Barnes-Farrell, 2024) and an Organizational Climate Scale for Public Service (Matheis et al., 2024). However, to the best of our knowledge, no research has been conducted applying the complete LLWI measure so far.

One reason for this lack may be that the 80-item LLWI is a comprehensive instrument, making it also time- and cost-intensive. As a result, the authors who developed the LLWI have advocated developing a short form (Wilckens et al., 2021) to improve the time and cost efficiency of its application. This is also highly beneficial in research, as a short form would alleviate participant burden, potentially increasing participation rates and data quality (Eisele et al., 2022; Galesic & Bosnjak, 2009; Kruyen et al., 2013). As such, a short form could facilitate in particular longitudinal or multiwave research (van der Heijden et al., 2018), which has so far only been sparsely conducted in this research field. Moreover, even though prior validation studies in Germany and the United States (Finsel et al., 2023c; Wilckens et al., 2021) have shown the overall nine-domain factor structure fit to be acceptable, an LLWI short form that reduces the complexity of the LLWI could further improve the overall model fit. Together with the shortened time to administer this assessment, this could increase the feasibility of using a comprehensive measure and not just selected subscales.

Reduced administration time is also beneficial in organizational contexts, where working time is valuable (Schaufeli et al., 2019). A shorter questionnaire could also be more easily integrated into regular employee surveys. Additionally, an LLWI short form could serve as an effective preliminary self-screening tool for organizations, highlighting areas for further in-depth analysis.

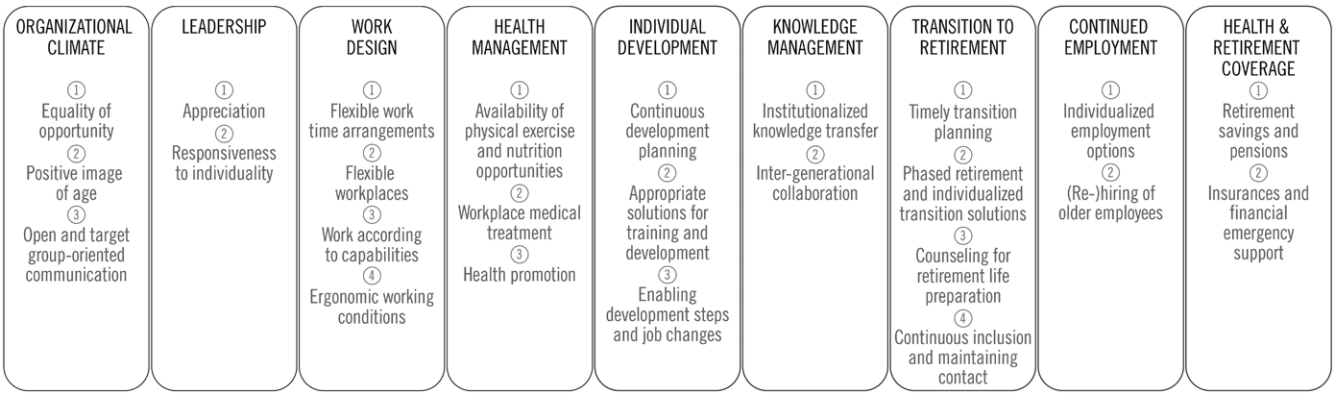

Figure 1. *Overview of the Later Life Workplace Index (LLWI).* Adapted from Wilckens et al. (2021).

Furthermore, there is as yet little empirical research on potential cross-national differences in the availability of organizational practices for older workers. The model of culture fit (Aycan et al., 1999) suggests that the availability of such practices is influenced by formal (e.g., legal standards) and informal aspects (e.g., specific cultural domains). According to this model, it is likely that cross-national differences in the availability of organizational practices would exist if the national contexts are dissimilar. Indeed, previous research on age-related HR practices has shown that institutional pressures can affect employer strategies toward older workers (e.g., Ollier-Malaterre et al., 2013). Likewise, studies on organizational practices in general demonstrate that cultural context can also have an impact (e.g., Aycan et al., 2000). This highlights the need to provide more language-specific versions of the LLWI to enable such comparisons. However, a prerequisite for cross-cultural mean difference tests is that test scores can be interpreted similarly across the countries being studied (van de Vijver & Tanzer, 2004). While actual differences in the availability of HR practices caused by formal and informal country-level factors could lead to different test scores in the LLWI-SF assessment across countries, cross-cultural differences in psychological meaning and perceptions could also affect the measurement. For example, the meaning of "retirement age" could change depending on whether mandatory or conventional retirement ages are in place. Similarly, there are slight differences in the meaning of terms like "aging" and "successful aging" across countries (Reich et al., 2020). Neglecting to examine such potential influences could cause bias when conducting cross-cultural comparisons (van de Vijver & Tanzer, 2004). Consequently, a cross-cultural validation of the LLWI-SF is needed to enable meaningful cross-cultural research on organizational practices for older workers. Compared to single-country studies, the development of a uniform short measure across countries significantly enhances the ability to conduct such cross-cultural research.

Therefore, this study aims to develop and validate a short form of the LLWI measure (LLWI-SF) in several countries to facilitate a more efficient assessment and cross-national comparisons of organizational practices for older workers. To determine whether the measurement model is applicable across different national contexts and languages, we have chosen countries that diverge along a number of dimensions in order to test its robustness and suitability for cross-national comparisons. Accordingly, the ten countries included in this study cover different degrees of various macro characteristics like culture (e.g., collectivistic vs. individualistic), older persons' employment rate (from 55% in Italy to 78% in Japan for people aged 55 to 64 in 2022; OECD, 2023), (mandatory) retirement ages, and share of the working-age population.

The Later Life Workplace Index

The LLWI was created by adopting a multistudy procedure. First, an initial qualitative eight-domain framework was developed by Wöhrmann et al. (2018) based on qualitative expert interviews in Germany. A ninth domain (health and retirement coverage) was added when Wilckens et al. (2020) integrated U.S. data into the model. Wilckens et al. (2021) then developed and cross-validated the original German 80-item measure in several consecutive quantitative studies among German employees and managers (Wilckens et al. 2021).

Each of the nine LLWI domains features two to four indicators (see gray sections in Figure 1), which are measured by two to four items each.[1] For example, the organizational climate domain entails equality of opportunity regardless of age, having a positive attitude towards older workers, and communicating a differentiated image of age within and outside the organization. The leadership domain includes managers showing appreciation toward their older workers and responding to their individual needs and abilities. Work design refers to flexible work time and workplace arrangements, adapting the work task to capabilities, and ensuring an ergonomic workplace design. The health management domain includes organizational practices aimed at maintaining and fostering older workers' health. Individual development refers to continuous development planning even in later life, development and job change opportunities, as well as target-group oriented training. The knowledge management domain describes formal and informal organizational practices that foster knowledge conservation and exchange between different age groups. Transition to retirement includes timely transition planning, providing individualized employment solutions for the transition, preparing older workers for retirement, and maintaining contact with retired workers. The continued employment domain refers to individualized employment options in retirement and the recruitment of older workers. Finally, the health and retirement domain entails organizational practices covering the information on and provision of retirement savings as well as supplemental health-related insurance (for detailed definitions, see Wilckens et al., 2021). Organizational practices relevant to older workers constitute a multidimensional construct (Boehm et al., 2021). Accordingly, the LLWI items cover age-specific (e.g., counseling for retirement life), age-inclusive (e.g., continuous development planning), and age-neutral practices (e.g., flexible work time arrangements).

The previous validation studies (Finsel et al., 2023c; Wilckens et al., 2021) provide insight into how the LLWI relates to associated constructs. First, the LLWI demonstrates convergent validity, as it is moderately to strongly related to two constructs measuring closely related organizational practices (i.e., age-inclusive HR practices and age-diversity climate; Boehm et al., 2014). Second, the LLWI shows only a small correlation with older workers' affect, demonstrating its discriminant validity as older employees' general mood should only show a small relation to relatively stable organizational-level practices (Wilckens et al., 2021). Third, concerning its criterion validity for older employees' outcomes, the LLWI is moderately related to attitudinal constructs such as work engagement, job satisfaction, and affective commitment. Moreover, it shows small to moderate correlations with health-related outcomes like perceived health, well-being, and work ability. The relationships with behavioral outcomes such as turnover intention and post-retirement work intention are lower than those for attitudinal and health-related workers' outcomes, but they are still significant. Only with performance not all nine LLWI domains show a significant correlation. Additionally, depending on the criterion under consideration, some LLWI

[1]To keep in line with the original authors (Wilckens et al., 2021), we use the word "indicator" as a conceptual term (i.e., subdomain). We are not using it in the statistical sense (i.e., item).

domains are more strongly correlated than others. Regarding health outcomes, for example, the domains work design and leadership show a stronger correlation than continued employment. As Finsel et al. (2023b) point out, various theories from the HR management, lifespan, motivational psychology, social psychology, vocational psychology, and economic literature can be applied to explain why the different types of organizational practices measured by the LLWI are related to diverse individual outcomes and how these relationships differ in strength for individual LLWI domains and outcomes considered. Since we aim to preserve the content validity of the LLWI in the short form, we expect comparable relationships for the LLWI-SF in our study (see Figure 2 for an overview). These hypothesized relationships are in line with common research guidelines recommending that convergent measures should demonstrate a strong ($r > .50$), discriminant measures a weak ($r < .30$), and criterion measures a theoretically grounded relationship to the scale under investigation (Cohen, 1988; Kline, 2005).

Materials and methods

Item selection

The development of the LLWI-SF was guided by several aims. Overall, we wanted to significantly reduce the length of the original 80-item measure. We also sought to preserve the conceptual model of the LLWI by keeping all nine domains as separate factors in the LLWI-SF. This would maintain the LLWI's scope of application since researchers could continue to use selected domain-level scales without having to administer the LLWI-SF as a whole.

To reduce the complexity while preserving content coverage as much as possible, we decided to retain the domain-level factor structure but drop the indicator level (Rammstedt & Beierlein, 2014). We further sought to preserve the instrument's psychometric properties by achieving acceptable internal consistencies for the nine domain-level scales and an acceptable model fit of the overall nine-domain factor structure. To ensure both the content validity and the reliability of the domain-level scales, we planned to retain at least one item per indicator and a minimum of three items per domain-level scale (Liden et al., 2015). Ultimately, we aimed to develop a cross-nationally structurally equivalent instrument that is universally applicable to diverse regulatory and cultural backgrounds, facilitating cross-national comparative research.

Several authors recommend developing a short form measure by combining qualitative and quantitative strategies to select items from a long measurement instrument (e.g., Kruyen et al., 2013; Stanton et al., 2002). We followed a five-phase strategy (see Figure 3 for an overview of the process). First, following the method proposed by Kruyen et al. (2013) and Smith et al. (2000), we started the item selection process with a qualitative strategy to ensure the preservation of the model content. An expert focus group was created that consisted of two of the authors of the German- and English-language versions of the LLWI, as well as five of the authors of this study, who had different national backgrounds and had prior knowledge of the LLWI measure. In line with standard practices regarding short-scale development (Kruyen et al.,

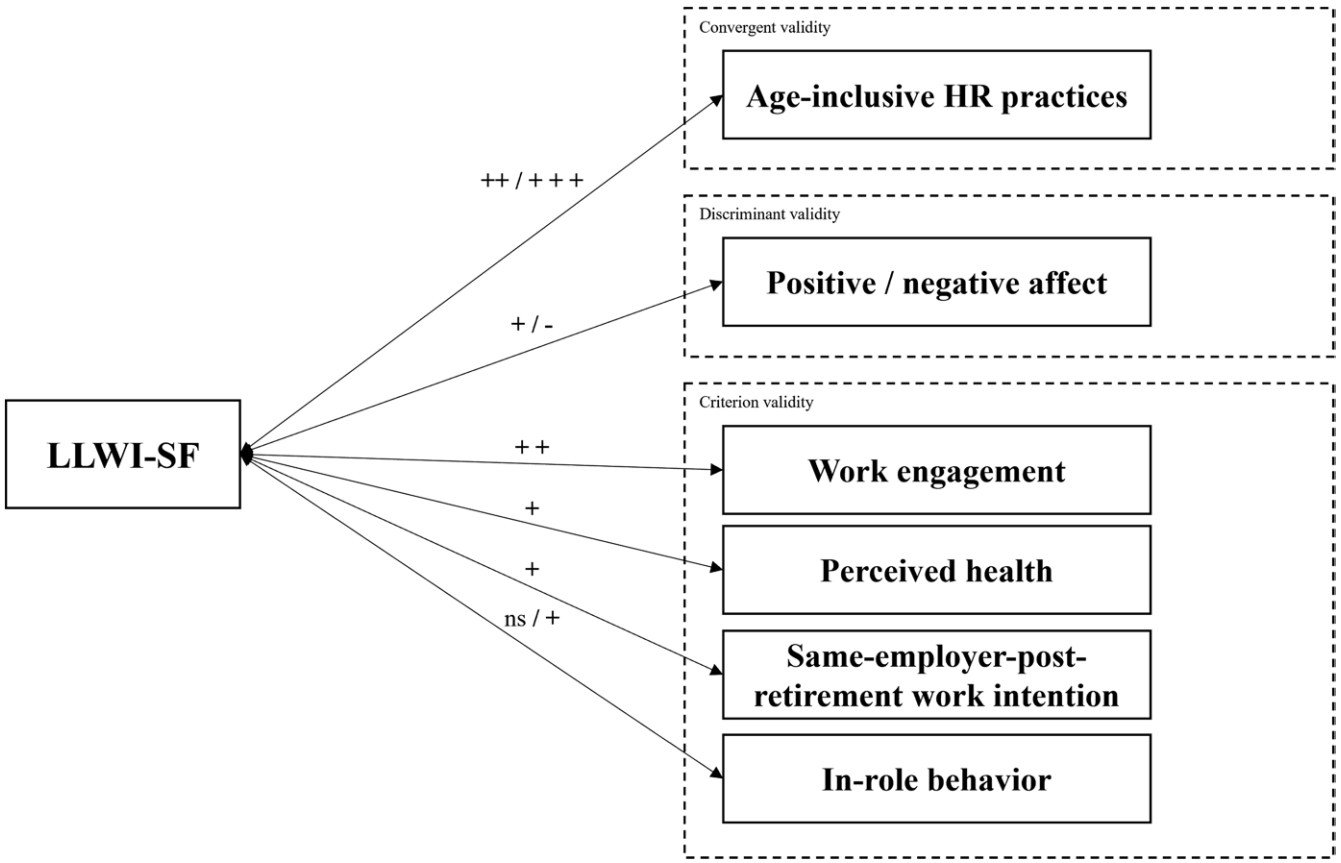

Figure 2. *Expected relationships with the validation scales.*

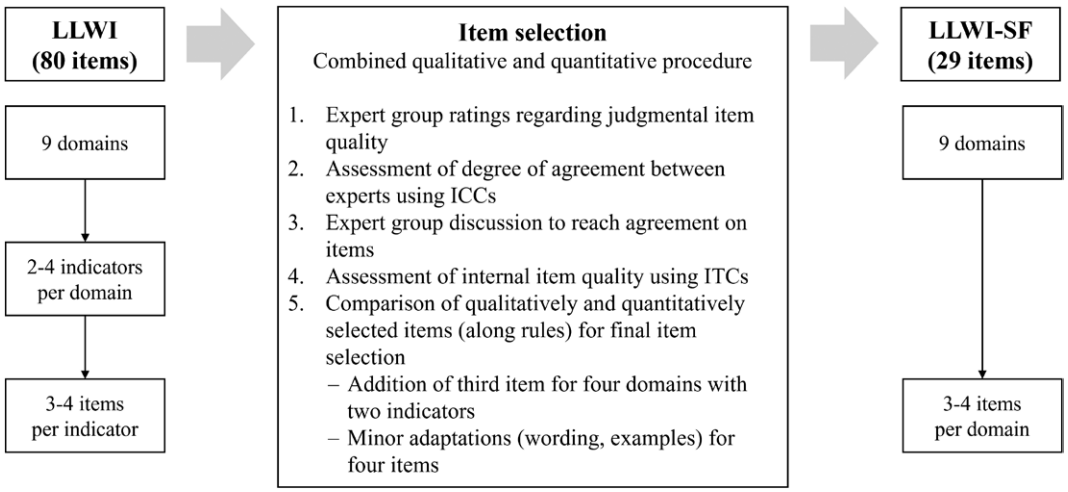

Figure 3. *Development process of the LLWI-SF.*

2013), each expert was asked to order the items of each domain regarding their perceived relevance for the respective indicator and domain construct as an assessment of judgmental item quality (Stanton et al., 2002).

In the second phase, since one item was to be selected from each of the 25 indicators, the degree of agreement between the experts (i.e., inter-rater reliability) regarding the item rating was assessed for each indicator using the intra-class correlation coefficient (ICC). The ICC was calculated using two-way random-effects models with multiple raters, focusing on absolute agreement (Koo & Li, 2016). If there was only poor or moderate inter-rater reliability (ICC < .75; Koo & Li, 2016) or if there was a tie between two or more items, the content of the items and the applicability across countries was discussed within the expert group. Two additional authors of the German-language LLWI participated in these discussions.

In the third phase, the expert focus group made a preliminary choice of one item for each indicator, which had been interpreted to best cover the spectrum of the indicator and domain construct (the most general item) and that was deemed to apply to most organizations and countries.

In the fourth phase, we combined these judgmental item quality results with the quantitative strategy by obtaining access to the data from the German and English LLWI validation studies (Finsel et al., 2023c; Wilckens et al., 2021). We used this data to individually calculate the corrected item-total-correlations (ITCs) on the domain level for the two German samples and the U.S. sample to assess internal item quality (Stanton et al., 2002). Using statistical criteria alone, the items with the highest ITC should be selected for the short form (Kruyen et al., 2013). However, in many cases, the items with the highest ITCs differed slightly between the three samples.

The items selected for each indicator based on the judgmental and internal item quality were then compared in the last phase. If different items were deemed to fit best, the expert focus group considered both criteria and made a final decision in accordance with our study objectives. In case of uncertainty, content coverage and cross-cultural applicability were prioritized compared to the highest ITCs. The reason for this prioritizing was that a statistics-driven item selection can lead to an optimization of reliability at the cost of a reduction of content validity (Thompson, 2007). With such a broad construct as the LLWI, item reduction is often inevitably related to a loss in internal consistency, since the remaining items must be rather heterogeneous to still cover the breadth of the construct (Rammstedt & Beierlein, 2014). In the four cases where the domain consisted of only two indicators (i.e., leadership, knowledge management, continued employment, and health and retirement coverage), an additional third item was selected for the scale based on the expert focus group ratings and the ITCs. The final version of the LLWI-SF comprises 29 items.

In two cases, the expert focus group made small changes to the original items to reflect the indicators' construct adequately. For one item of the indicator flexible work time arrangements, the item wording was changed from "Employees of our organization have enough flexibility in their working time organization to appropriately address *unforeseen events in their private lives*" to "Employees of our organization have enough flexibility in their working time organization to appropriately address *their personal needs*" in alignment with the item wording of the other three items of the indicator that were not selected for the short form. For one item of the indicator continuous development planning, the item wording was changed from "In our organization, development prospects and qualification requirements are identified *for* employees, regardless of age" to "In our organization, *individual* development prospects and qualification requirements are identified *with* employees, regardless of age" to account for the missing information from the removed items that highlights that employees are actively included in the development planning. In two more cases (LLWI-SF items HM-2 and RC-1 in Table A1), an additional example was added to the item to increase comprehensibility across countries. The English-language items are displayed in Table A1. All other language versions of the LLWI-SF can be obtained from Table S1 in the Supplemental Online Material (https://doi.org/10.48548/pubdata-1716).

Sample

Ten national samples were included in the current study (N = 3,424). All data was collected cross-sectionally between October 2022 and June 2023 using panel providers or survey platforms (i.e., Bilendi, Norstat, Remember, TGM Research, Toluna, and Webankieta) to increase the representativity of the

samples. All samples were collected online except the Polish sample, which was collected both online and via a paper-pencil survey. Several screening criteria were employed. First, all participants were required to live in their respective country and to have the respective language as their mother tongue to ensure the comprehensibility of the items. Second, participants had to be working at an organization with at least 30 employees for a minimum of 1 year and with a minimum of 20 hr per week to ensure that they had access to and knowledge of organizational practices for older workers. Third, participants in all samples were required to be at least 50 years old to be included in the target group of the practices covered in the LLWI.

The data screening procedure differed slightly between the participating countries. Except for Norway, all countries applied instructed response items and survey timings to detect and remove inattentive participants. Six countries (i.e., Belgium, Germany, Japan, Portugal, the Netherlands, and the United States) also applied the LongString Index to screen out inattentive participants. The national sample sizes ranged from 140 to 444. Detailed information regarding the sample compositions can be obtained from Table 1.

Measures

First, for the LLWI-SF, a translation-back translation technique (Brislin, 1970) was used to translate the items from German or English into other languages. If necessary for the cultural adaptation of the LLWI items, small adaptations were made to the item wording (e.g., providing a different example). All changes were made in consultation with the original authors of the LLWI to ensure the consistency of the item content. Second, concerning the measures for testing convergent, discriminant, and criterion validity used in this study, we selected measures that had been previously used in the LLWI validation studies (Finsel et al., 2023c; Wilckens et al., 2021) to compare the strengths of the relationships. If validated translations were unavailable, a translation-back translation technique was applied.

LLWI-SF

The LLWI-SF was administered in all ten countries in their respective language versions (see Table A1 for the English-language items).[2] The responses were measured on a rating scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).

Measure for testing convergent validity

The age-inclusive HR practices items developed by Boehm et al. (2014) were used to test convergent validity, since the items of this scale cover several of the LLWI domain constructs, and indeed, the scale appeared to be strongly correlated with all LLWI domains (Finsel et al., 2023c; Wilckens et al., 2021). The five items were measured on a 5-point Likert scale ranging from 1 (*very low intensity*) to 5 (*very high intensity*). Cronbach's alpha ranged between .81 and .92 (see Supplemental Online Material, https://doi.org/10.48548/pubdata-1716).

Measure for testing discriminant validity

In line with the previous validation studies for the LLWI, the Positive and Negative Affect Schedule Short Form (PANAS-SF; Thompson, 2007) was used to test for discriminant validity. The ten items were measured on a 5-point Likert scale ranging from 1 (*never*) to 5 (*always*). For most samples, Cronbach's alpha ranged between .57 and .89 for the positive and negative affect subscales. In Norway, however, internal consistency was only .41 for the positive affect subscale. We removed one item with a low ITC of .02, increasing internal consistency to .60.

Measures for testing criterion validity

To test criterion validity, we aimed to cover attitudinal, health-related, and behavioral criteria. To enable comparability, we selected outcomes for older workers that were used in the previous validation studies of the LLWI (Finsel et al., 2023c; Wilckens et al., 2021) to cover these three categories. In particular, we used work engagement, health, and same-employer-post-retirement work intention, as these outcomes are highly relevant for keeping older workers motivated and healthy enough to continue working in later life (de Wind et al., 2014; Fisher et al., 2016). Work engagement was measured using the three-item short measure developed by Schaufeli et al. (UWES-3; Schaufeli et al., 2019). The items were measured on a scale from 1 (*never*) to 7 (*always*). We further administered four items to measure perceived health (Adams & Beehr, 1998) and three items to measure the same-employer-post-retirement work intention (Wöhrmann et al., 2013). Lastly, to build upon and advance the study conducted by Finsel et al. (2023c), who measured peer-rated performance and were only able to demonstrate significant correlations with some LLWI domains, we also included four items to measure self-rated in-role behavior (Eisenberger et al., 2001). All items were measured on a 7-point Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*).[3] Cronbach's alpha was above .77 for all scales in all subsamples.

Data analysis

Data analyses were conducted in R version 4.3.0 and Mplus version 8.8. As a first step, we calculated the corrected ITCs, means, standard deviations, skewness, and kurtosis for the LLWI-SF items per country. Next, we computed the means, standard deviations, internal consistencies, and intercorrelations of the nine LLWI-SF subscales per country and the overall sample. To check the factorial validity, we conducted confirmatory factor analyses (CFA) using a robust maximum likelihood estimation based on Yuan-Bentler, since the data was nonnormally distributed (Yuan & Bentler, 2007). To account for missing data, we applied the full information maximum likelihood estimation (Newman & Cottrell, 2015).[4] The Kaiser–Meyer–Olkin measure of sampling adequacy and Bartlett's test of sphericity both indicated that the data was suitable for CFA. Following standard CFA practices (see Jackson et al., 2009), we used several fit indices to assess our model fit. The robust comparative fit index (CFI) and Tucker–Lewis index (TLI) should be higher than .90 to indicate an acceptable fit (McDonald & Ho, 2002). Moreover, the

[2]Table S1 in the Supplemental Online Material (https://doi.org/10.48548/pubdata-1716) contains the LLWI-SF items in Dutch, Flemish, German, Italian, Japanese, Korean, Norwegian, Polish, and Portuguese language.

[3]There were two exceptions to this procedure. First, in the Italian sample, the same-employer-post-retirement work intention was measured on a 5-point Likert scale ranging from 1 (*strongly disagree)* to 5 (*strongly agree)*. Second, in the Korean sample, we only had data on the work engagement criterion.

[4]For scale building and correlation analysis, we handled item-level missingness following the recommendations by Newman (2014) and computed scale scores from the mean of the available items. For the correlation tables, pairwise deletion was used to handle construct-level missingness.

Table 1. Sample characteristics.

	Belgium	Germany	Italy	Japan	Korea
N	444	387	408	349	350
Mean age (*SD*)	57.0 (4.3)	57.2 (4.8)	57.1 (4.2)	57.4 (5.3)	55.3 (4.5)
% Female	35.8%	45.5%	56.4%	16.3%	23.7%
Educational level	Primary level 1.1% Secondary level 30.9% Higher level 68.0%	Primary level .0% Secondary level 70.5% Higher level 29.5%	Primary level 1.7% Secondary level 56.6% Higher level 41.7%	Primary level 1.1% Secondary level 32.7% Higher level 66.2%	Primary or secondary level 10.0% Higher level 90.0%
% Management jobs	25.2%	33.9%	46.6%	33.8%	85.7%
Tenure	1–6 years 13.5% >6 years 86.5%	1–6 years 15.8% >6 years 84.2%	>1 year 100.0%	1–6 years 20.0% >6 years 80.0%	1–6 years 32.0% >6 years 68.0%
More than half the working time spent doing office work	57.0%	60.5%	75.2%	62.5%	–
More than half the working time spent doing physical work	26.5%	26.4%	18.1%	17.2%	–
Organization size[a]	Small 32.7% Medium 27.5% Large 39.9%	Small 42.1% Medium 28.2% Large 29.7%	Small 16.7% Medium 27.0% Large 56.4%	Small 43.9% Medium 30.9% Large 25.2%	Small 44.0% Medium 24.3% Large 31.7%
Sector	Service 30.2% Industry 28.6% Education 11.5% Retail 5.2% Public 23.9% Other .7%	Service 40.3% Industry 34.1% Education 3.9% Retail 9.6% Public 11.6% Other .5%	Service 38.9% Industry 16.4% Education 17.9% Retail 5.1% Public 20.3% Other 1.7%	Service 38.4% Industry 39.8% Education 6.3% Retail 9.2% Public 6.0% Other .3%	Service 40.0% Industry 44.3% Education 8.0% Retail 7.7% Public .0%
	Norway	**Poland**	**Portugal**	**The Netherlands**	**United States**
N	140	353	306	317	370
Mean age (*SD*)	58.7 (4.9)	55.0 (4.4)	55.1 (4.5)	57.1 (5.0)	58.4 (6.1)
% Female	52.1%	53.5%	42.5%	32.8%	59.2%
Educational level	Primary level 3.6% Secondary level 35.0% Higher level 61.4%	Primary level 14.4% Secondary level 28.3% Higher level 57.2%	Primary level 15.4% Secondary level 30.4% Higher level 54.2%	Primary level .0% Secondary level 56.1% Higher level 43.8%	Primary level .8% Secondary level 40.0% Higher level 59.1%
% Management jobs	23.6%	38.2%	39.2%	42.3%	44.1%
Tenure	1–6 years 15.7% >6 years 84.3%	1–6 years 27.5% >6 years 72.2%	1–6 years 14.7% >6 years 85.3%	1–6 years 23.3% >6 years 76.7%	1–6 years 33.3% >6 years 66.7%
More than half the working time spent doing office work	58.6%	49.3%	61.1%	42.9%	56.8%
More than half the working time spent doing physical work	17.1%	32.6%	23.2%	34.7%	28.4%
Organization size[a]	Small 46.4% Medium 9.3% Large 44.3%	Small 70.8% Medium 17.3% Large 11.9%	Small 40.2% Medium 33.7% Large 26.2%	Small 41.6% Medium 29.0% Large 29.4%	Small 41.6% Medium 28.9% Large 29.5%
Sector	Service 42.1% Industry 24.3% Education 14.3% Retail 2.9% Public 10.7% NA 5.7%	Service 32.3% Industry 34.0% Education 16.4% Retail 7.1% Public 10.2%	Service 26.8% Industry 26.8% Education 17.0% Retail 7.5% Public 21.9%	Service 36.0% Industry 30.3% Education 7.6% Retail 11.4% Public 13.6% Other 1.3%	Service 47.6% Industry 24.6% Education 11.4% Retail 10.3% Public 5.4% Other .8%

[a]Small organizations 30–499 employees, medium-sized organizations 500–4,999 employees, large organizations more than 5,000 employees. Exception Italy: Small organizations 30–49 employees, medium-sized organizations 50–249 employees, large organizations 250 and more employees. Exception Korea: Small organizations 30–299 employees, medium-sized organizations 300–1,000 employees, large organizations more than 1,000 employees.

robust root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR) should be below .08 to indicate an acceptable fit (Browne & Cudeck, 1992; Hu & Bentler, 1999).

We used multigroup CFAs to test for measurement invariance. First, we tested for configural invariance (i.e., that the basic factor structure is invariant across countries; Putnick & Bornstein, 2016). Configural invariance was supported if the fit

Table 2. Confirmatory factor analysis results for the measurement model.

Models	χ^2	*df*	CFI	RMSEA	SRMR
1-factor	35,916.858***	1,080	.574	.097	.097
2-factor	30,89.027***	1,079	.636	.090	.121
3-factor	29,294.361***	1,077	.655	.087	.120
6-factor	15,347.329***	1,065	.825	.063	.056
14-factor (proposed model)	5,685.832***	989	.943	.037	.039

Note. N = 3,424. 1-factor: all variables load on a single factor. 2-factor: LLWI-SF and age-inclusive HR practices items load on a single factor, while the criteria items load on the second factor. 3-factor: LLWI-SF items load on a single factor, age-inclusive HR practices items load on the second factor, criteria items load on the third factor. 6-factor: LLWI-SF items load on a single factor, while age-inclusive HR practices and criteria items all load on their corresponding factors. 14-factor: each item loads on its corresponding first-order factor.
***$p < .001$.

indices of the baseline model with freely estimated parameters indicated an acceptable model fit. If configural invariance was supported, we tested for metric invariance (i.e., that the items load similarly onto the factors across countries; Putnick & Bornstein, 2016). To do this, we constrained the factor loadings to be equal across the countries. Following Chen's (2007) recommendations, metric invariance was supported if the change in CFI was <–.010, the change in RMSEA <.015, and the change in SRMR <.030 between the baseline and the metric model. If metric invariance was supported, we tested for scalar invariance (i.e., that the item intercepts are similar across countries) by constraining the item intercepts to be equal. Following Chen's (2007) recommendations again, scalar invariance was supported if the change in CFI was <–.010, the change in RMSEA <.015, and the change in SRMR <.010. If the model fit change exceeded the cutoffs, this indicated that at least one item intercept differed across the countries. In that case, we tested for partial scalar invariance by sequentially releasing item intercept constraints with the highest modification indices within countries until the change in model fit met the cutoffs (Putnick & Bornstein, 2016). Partial scalar invariance was then supported if at least half of the items of a subscale were invariant (Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). If (partial) scalar invariance was supported, we tested for residual invariance (i.e., that the item residuals are similar across countries; Putnick & Bornstein, 2016). In the case of partial scalar invariance, the freed parameters were retained in this model (Putnick & Bornstein, 2016). The change in model fit was then interpreted by applying the same cutoff values used for scalar invariance (Chen, 2007). If the change exceeded these cutoffs, we tested for partial residual invariance again by sequentially releasing item residual constraints. In the final model, we restricted the factor means to zero across all groups to test for mean invariance (Vandenberg & Lance, 2000). If the model fit decreases, this indicates that factor means are not equal across countries. This step is possible even without reaching (partial) residual invariance as long as (partial) scalar invariance is supported (Vandenberg & Lance, 2000). We did not use the reported Satorra-Bentler scaled χ^2 difference tests to assess measurement invariance since prior research has shown it to be sensitive to large sample sizes and nonnormality in the data, making it problematic for measurement invariance testing (Chen, 2007; Meade et al., 2008; Steinmetz, 2013).

We calculated the Pearson correlation coefficient on the domain level to test for convergent, discriminant, and criterion validity. To establish predictive and incremental validity, we performed hierarchical linear regression analyses for each country using full information maximum likelihood to handle construct-level missingness. In the first step, the criterion variables were regressed on the convergent variable age-inclusive HR practices. In a second step, the nine LLWI-SF domains were added as predictors. Adjusted R^2 are reported to account for the high number of predictors in these regressions.

Since participants might differ in their evaluation of organizational practices based on their sociodemographic characteristics (e.g., age; Nishii et al., 2018) or their organization's characteristics (e.g., industry type; Wilckens et al., 2021), we also took a closer look at the relation of the LLWI-SF domains with participants' gender and work type as well as their organization's size and industry type. For this purpose, we created dummy variables if necessary.

Measurement model

We performed a set of alternative CFAs to assess the construct validity of the variables measured in the study (i.e., LLWI-SF domains, age-inclusive HR practices, work engagement, perceived health, post-retirement work intention, in-role behavior) and compared the fit of the hypothesized measurement model to alternative models. The measurement model where each of LLWI-SF items load on the corresponding domains showed an adequate fit to the data ($\chi 2$ = 5,685.833, df = 989, $p < .001$, CFI = .94, RMSEA = .04, SRMR = .04) and fit better than all potential alternative models (see Table 2). These results indicate that the measurements utilized in this study are empirically distinct.[5]

Results

Item statistics, intercorrelations, and internal consistency

The ITCs per country are displayed in Table 3. The item means, standard deviations, skewness, and kurtosis per country are included in the Supplemental Online Material (https://doi.org/10.48548/pubdata-1716). All items had a

[5]To test the potential for common method variance further, we applied a correlational marker approach. In particular, we calculated partial correlations between the LLWI-SF subscales and the four criterion variables controlling for positive and negative affect (Simmering et al., 2015). The average decrease in effect size was only .06 (Δr = [.01, .14]), indicating that common method bias only had a small effect.

Table 3. Corrected item-total-correlations for LLWI-SF items across countries on domain level.

Item	Belgium	Germany	Italy	Japan	Korea	Norway	Poland	Portugal	The Netherlands	United States
OC-1	.66	.62	.54	.60	.67	.62	.61	.59	.52	.62
OC-2	.48	.70	.60	.69	.72	.74	.62	.69	.64	.67
OC-3	.71	.50	.40	.61	.64	.53	.44	.56	.47	.48
LE-1	.85	.86	.83	.79	.77	.74	.72	.77	.77	.84
LE-2	.89	.86	.89	.85	.77	.83	.82	.85	.79	.87
LE-3	.84	.86	.85	.77	.75	.84	.74	.83	.82	.81
WD-1	.58	.66	.54	.70	.60	.72	.58	.54	.62	.57
WD-2	.70	.65	.74	.71	.75	.74	.61	.59	.60	.70
WD-3	.44	.45	.48	.59	.75	.62	.47	.46	.31	.53
WD-4	.50	.52	.48	.61	.76	.56	.52	.48	.52	.65
HM-1	.58	.62	.63	.42	.66	.59	.55	.50	.59	.58
HM-2	.55	.73	.62	.70	.76	.63	.68	.60	.59	.76
HM-3	.72	.79	.61	.69	.73	.77	.57	.59	.66	.65
ID-1	.59	.73	.70	.74	.77	.76	.65	.71	.51	.58
ID-2	.67	.70	.68	.77	.76	.77	.61	.71	.57	.64
ID-3	.55	.68	.68	.72	.75	.70	.56	.65	.27	.64
KM-1	.65	.76	.72	.80	.80	.64	.69	.66	.53	.63
KM-2	.68	.75	.79	.77	.80	.70	.78	.71	.57	.74
KM-3	.77	.78	.84	.77	.80	.76	.75	.77	.60	.72
TR-1	.72	.85	.80	.81	.69	.57	.70	.78	.70	.86
TR-2	.74	.82	.83	.71	.62	.68	.67	.73	.54	.84
TR-3	.77	.84	.81	.85	.83	.59	.72	.75	.67	.83
TR-4	.58	.76	.67	.73	.74	.59	.58	.59	.54	.80
CE-1	.53	.71	.65	.55	.66	.59	.51	.58	.56	.63
CE-2	.57	.71	.60	.63	.68	.69	.59	.53	.59	.67
CE-3	.38	.46	.35	.43	.45	.59	.55	.34	.19	.44
RC-1	.76	.56	.71	.70	.77	.70	.70	.74	.67	.70
RC-2	.62	.72	.67	.71	.79	.57	.74	.73	.74	.61
RC-3	.67	.65	.57	.67	.65	.46	.73	.54	.65	.67

Note. Corrected item-total-correlations between the single item and the other items of the domain, not including the respective single item, are reported.

skewness and kurtosis below the accepted limits of ±2.0 and ±7.0 (Curran et al., 1996), showing that we did not have extreme nonnormally distributed data. Generally, the LLWI-SF items had an ITC above the accepted cutoff of ≥.3 (Nunnally & Bernstein, 1994), indicating that items correlated well with the domain-level scales. Only in the Dutch sample were the ITCs for the items ID-3 and CE-3 below the cutoff (.27 and .19). In light of cross-country comparisons and the aim to safeguard construct validity, we keep the respective items in the LLWI-SF subscales for this study.

The domain-level means, standard deviations, and internal consistencies are displayed in Table 4. Internal consistency for most subscales was acceptable on the country level (Nunnally & Bernstein, 1994). In 38 cases (42%), Cronbach's alpha ranged between .70 and .80. In 40 cases (44%), it ranged between .81 and .90. In seven cases (8%), Cronbach's alpha even exceeded .90. Only for one case in the Belgian, Italian, and Portuguese samples and for two cases in the Dutch sample (6%) was Cronbach's alpha of the shortened subscales below < .70. These five cases with moderate reliability of .62–.69 refer to the subscales *organizational climate*, *individual development*, and *continued employment*. This is a common problem that can occur when the development process of the short scale is not just statistically driven but also follows a qualitative procedure to ensure its content validity (Ziegler et al., 2014). Looking more closely at the affected subscales and samples, the inter-item correlations ranged between .31 and .56 in the Belgian, .33 and .58 in the Italian, .27 and .60 in the Portuguese, and .15 and .64 in the Dutch sample, indicating that the items are related enough to compute the scale despite the lower internal consistencies (Briggs & Cheek, 1986). Regarding the overall sample, Cronbach's alpha ranged between .70 and .80 for the subscales *organizational climate*, *work design*, *health management*, and *continued employment*. It ranged between .81 and .90 for the subscales *individual development*, *knowledge management*, *transition to retirement*, and *health and retirement coverage*. Finally, it exceeded .90 for the subscale *leadership*. Taking the highest internal consistency reported for each LLWI subscale in Wilckens et al. (2021) and Finsel et al. (2023c), the loss in reliability for the overall sample ranged between .03 and .17 (.10 on average). While the internal consistencies were lower, it is quite common for short scales to have lower reliabilities compared to their longer counterparts due to the fewer number of items (Kruyen et al., 2013).

Table 4. Means, standard deviations, and internal consistencies across countries.

Subscale	Belgium	Germany	Italy	Japan	Korea	Norway	Poland	Portugal	The Netherlands	United States	All
	M (*SD*, *α*)	*M* (*SD*, *α*)	*M* (*SD*, *α*)	*M* (*SD*, *α*)	*M* (*SD*, *α*)	*M* (*SD*, *α*)	*M* (*SD*, *α*)	*M* (*SD*, *α*)	*M* (*SD*, *α*)	*M* (*SD*, *α*)	*M* (*SD*, *α*)
OC	4.87 (1.33, .81)	5.25 (1.23, .77)	4.45 (1.38, .69)	4.55 (1.32, .79)	4.92 (1.21, .82)	5.42 (1.09, .78)	4.98 (1.29, .73)	4.20 (1.45, .77)	4.82 (1.24, .71)	4.95 (1.31, .75)	4.81 (1.33, .76)
LE	4.75 (1.38, .93)	4.95 (1.48, .93)	4.29 (1.63, .93)	4.38 (1.34, .90)	4.77 (1.28, .88)	5.31 (1.25, .90)	4.80 (1.38, .88)	3.96 (1.54, .91)	4.89 (1.38, .90)	5.41 (1.37, .92)	4.72 (1.47, .92)
WD	4.20 (1.23, .76)	4.29 (1.31, .77)	3.48 (1.37, .76)	4.05 (1.23, .82)	4.16 (1.42, .86)	4.38 (1.39, .83)	4.05 (1.36, .75)	3.63 (1.24, .73)	4.30 (1.24, .72)	4.61 (1.27, .75)	4.08 (1.35, .78)
HM	3.72 (1.42, .78)	3.63 (1.66, .84)	3.01 (1.48, .78)	3.83 (1.31, .76)	4.49 (1.34, .85)	3.85 (1.53, .81)	3.49 (1.47, .76)	3.12 (1.47, .74)	4.12 (1.51, .78)	4.18 (1.54, .79)	3.75 (1.52, .78)
ID	4.08 (1.32, .77)	4.31 (1.56, .84)	3.63 (1.60, .83)	3.95 (1.33, .86)	4.44 (1.38, .87)	4.27 (1.41, .86)	4.14 (1.47, .77)	3.95 (1.43, .83)	4.57 (1.20, .63)	5.03 (1.32, .78)	4.23 (1.47, .82)
KM	3.95 (1.41, .84)	4.39 (1.66, .88)	3.78 (1.72, .89)	3.78 (1.43, .89)	4.39 (1.42, .90)	4.36 (1.35, .84)	4.64 (1.46, .86)	4.02 (1.51, .85)	4.69 (1.27, .74)	4.90 (1.42, .83)	4.27 (1.52, .86)
TR	2.99 (1.43, .85)	3.35 (1.79, .92)	2.75 (1.54, .90)	3.46 (1.44, .90)	4.10 (1.44, .87)	3.80 (1.43, .79)	3.67 (1.47, .84)	2.68 (1.44, .86)	3.65 (1.42, .80)	3.74 (1.68, .91)	3.38 (1.58, .88)
CE	3.54 (1.28, .68)	3.95 (1.57, .78)	3.26 (1.41, .71)	3.84 (1.30, .71)	4.19 (1.44, .76)	3.88 (1.40, .78)	4.11 (1.43, .73)	3.16 (1.28, .66)	4.49 (1.24, .62)	4.58 (1.41, .74)	3.89 (1.45, .71)
RC	3.47 (1.79, .82)	4.08 (1.74, .80)	3.02 (1.61, .80)	3.82 (1.43, .83)	4.05 (1.54, .86)	3.59 (1.72, .74)	3.82 (1.58, .85)	2.22 (1.46, .80)	3.86 (1.61, .83)	4.96 (1.52, .81)	3.71 (1.74, .83)

Note. We performed additional analysis to calculate the Omega total score for the subscales (Revelle, 2024). The scores were similar or slightly higher (0.00–0.05) than the Cronbach's alpha value of the respective subscale. OC = organizational climate; LE = leadership; WD = work design; HM = health management; ID = individual development; KM = knowledge management; TR = transition to retirement; CE = continued employment; RC = health and retirement coverage.

Table 5 displays the inter-domain correlations for the combined overall sample. The inter-domain correlations for the individual countries are included in the Supplemental Online Material (https://doi.org/10.48548/pubdata-1716). The inter-domain correlations were quite similar to those reported in Wilckens et al. (2021; average difference of .07) and Finsel et al. (2023c; average difference of .10). The highest correlations were found in the Korean sample, while the lowest correlations were identified for the Polish sample.

Confirmatory factor analysis

The results of the single-group confirmatory factor analysis for each country and the overall sample are displayed in Table 6. Generally, the results showed an acceptable to good fit with the data. The Norwegian sample was the sole sample that did not yield an acceptable fit based on the cutoff criteria we defined for the CFI and TLI (CFI = .88, TLI = .86). However, since the CFI is very close to the cutoff of .90 and as the subscale reliabilities were all above .70, we will continue with the proposed model in the Norwegian sample. Moreover, the analysis indicated that the covariance matrix was not positive definite for the Dutch, Norwegian, and U.S. samples. This was demonstrated by the model's determinant being 0 for all three samples. Looking at the covariance matrices, we found no negative variance that could cause this finding. Some of the covariances did, however, exceed a value of 1. The correlation matrices showed no correlations greater or equal to 1 among the latent variables, albeit some correlations exceeded .90. Hence, the reason for the nonpositive determinants might be a linear dependency between some of the subscales. In particular, there was a correlation of .94 and .90 between the subscales *individual development* and *knowledge management* in the Dutch and U.S. samples. Moreover, the correlation between the subscales *transition to retirement* and *continued employment* ranged between .91 and .98 in the three samples. Merging the subscales with correlations above .90 led to significantly worse fitting models for the Dutch and U.S. samples ($\Delta\chi^2 = 31.40$, $\Delta\, df = 15$, $p > .01$, Δ CFI = .006, Δ TLI = .003, Δ RMSEA = .001, Δ SRMR = .001 for the Netherlands; $\Delta\chi^2 = 83.95$, $\Delta\, df = 15$, $p > .001$, Δ CFI = .015, Δ TLI = .014, Δ RMSEA = .005, Δ SRMR = .007 for the United States). For the Norwegian sample, the model fit did not decrease significantly, but it did not increase either ($\Delta\chi^2 = 16.11$, $\Delta\, df = 8$, $p = .05$, Δ CFI = .007, Δ TLI = .005, Δ RMSEA = .001, Δ SRMR = .002). Furthermore, looking at the composite means for the Dutch, Norwegian, and U.S. samples, the inter-domain correlations did not exceed .75 and were comparable to those reported in Wilckens et al. (2021) and Finsel et al. (2023c). As a result, we will continue using the nine-factor solution to allow for measurement invariance testing since the model fit of the nine-factor solution was acceptable to good, and the correlations between the composite domain means indicate sufficient discriminant validity.

The LLWI-SF model fit for the overall sample ($\chi^2 = 2{,}526.51$, $df = 341$, CFI = .95, RMSEA = .05, SRMR = .03) is better than the model fits reported for the LLWI in Wilckens et al. (2021; $\chi^2 = 6{,}309.60$, $df = 3{,}024$, CFI = .91, RMSEA = .04, SRMR = .07 and $\chi^2 = 5{,}348.10$, $df = 3{,}024$, CFI = .89, RMSEA = .05, SRMR = .07) and Finsel et al. (2023c; $\chi^2 = 6{,}094.33$, $df = 3{,}024$, CFI = .83, SRMR = .09), indicating that the reduced complexity of the LLWI-SF is beneficial.

Table 5. Means, standard deviations, and correlations for the overall sample.

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9
1. Organizational climate	4.81	1.33	-								
2. Leadership	4.72	1.47	.67***	-							
3. Work design	4.08	1.35	.46***	.60***	-						
4. Health management	3.75	1.52	.39***	.52***	.63***	-					
5. Individual development	4.23	1.47	.51***	.66***	.64***	.65***	-				
6. Knowledge management	4.27	1.52	.53***	.67***	.59***	.60***	.73***	-			
7. Transition to retirement	3.38	1.58	.44***	.56***	.59***	.65***	.64***	.66***	-		
8. Continued employment	3.89	1.45	.50***	.56***	.58***	.56***	.63***	.63***	.72***	-	
9. Health and retirement coverage	3.71	1.74	.35***	.48***	.54***	.57***	.55***	.51***	.63***	.57***	-
10. Age-inclusive HR practices	3.05	0.95	.57***	.62***	.53***	.53***	.64***	.63***	.59***	.65***	.49***
11. Positive affect	3.70	0.65	.23***	.27***	.16***	.12***	.23***	.26***	.13***	.17***	.12***
12. Negative affect	2.06	0.65	–.14***	–.17***	–.10***	–.04*	–.11***	–.13***	–.03	–.07***	–.04*
13. Work engagement	4.82	1.31	.33***	.42***	.26***	.24***	.34***	.37***	.27***	.28***	.19***
14. Perceived health	4.95	1.30	.27***	.28***	.23***	.18***	.23***	.26***	.19***	.21***	.16***
15. Postretirement work intention	3.69	1.78	.22***	.31***	.31***	.29***	.30***	.28***	.32***	.34***	.27***
16. In-role behavior	6.25	0.81	.19***	.19***	.07***	.02	.13***	.15***	.01	.09***	.04*
17. Age	56.75	4.94	.03	.06***	.03*	.00	.03	.01	–.03	.02	.01
18. Gender[a]	0.42	0.49	.01	.00	–.07***	–.11***	–.04*	–.01	–.06***	–.02	–.04*
19. Work type: > 50% office	0.59	0.49	–.05**	–.03	–.04*	–.05**	–.02	–.05**	–.02	–.08***	–.01
20. Work type: > 50% manual	0.25	0.44	.05**	–.04*	–.09***	–.02	–.04*	.00	–.05**	.05*	–.04*
21. Employer size: 30–499	0.49	0.50	.06**	.01	–.11***	–.14***	–.09***	.00	–.02	.01	–.12***
22. Employer size: 500–4,999	0.23	0.42	–.03	–.02	.06**	.06**	.03	.00	–.01	.00	.04*
23. Employer size: >5,000	0.28	0.45	–.03	.01	.06***	.10***	.07***	.00	.04*	.00	.09***
24. Sector: Services	0.37	0.48	.04*	.07***	.11***	.08***	.07***	.06***	.04*	.04*	.10***
25. Sector: Industry	0.31	0.46	–.02	–.03	.00	.05**	–.01	–.01	.02	.01	.09***
26. Sector: Public	0.13	0.33	–.06**	–.08***	–.02	–.07***	–.06***	–.08***	–.08***	–.09***	–.17***

Note. N = 3,424. For clarity, only the correlations for the LLWI-SF subscales are displayed. A full correlation table can be obtained from the corresponding author upon request.
[a]1 = female, 0 = male.
*p < .05.
**p < .01.
***p < .001.

Measurement invariance

The results of the multi-group CFA are displayed in Table 7. The configural model demonstrated good fit to the data (χ^2 = 6,840.05, df = 3,410, p < .001, CFI = .931, RMSEA = .054, SRMR = .052), supporting configural invariance. Using the criteria of ΔCFI < .010, ΔRMSEA < .015, and ΔSRMR < .030, the change in model fit between the configural and metric models further indicated that full metric invariance was achieved (ΔCFI = .008, ΔRMSEA = .002, and ΔSRMR = .011).

Next, we applied the criteria of ΔCFI < .010, ΔRMSEA < .015, and ΔSRMR < .010 to the change in model fit between the metric and scalar models. Full scalar invariance was not supported (ΔCFI = .059, ΔRMSEA = .017, and ΔSRMR = .019). Guided by modification indices, we sequentially freed intercepts for specific items in individual countries to achieve partial scalar invariance. Significant differences in intercepts were observed for the ten items OC-3, LE-1, WD-2, WD-3, HM-2, ID-2, KM-1, TR-1, TR-2, CE-3, and RC-2. These items, such as WD-3, HM-2, ID-2, KM-1, and RC-2, largely reflect policies and procedures (e.g., task adjustments, therapeutic assistance, training, knowledge transfer, advice on financial security), whereas OC-3, LE-1, and CE-3 also reflect cultural considerations (e.g., openness about aging, appreciative leadership, hiring practices). However, even after releasing 52 individual item intercepts (3–7 per country), the change in model fit did not support partial scalar invariance (ΔCFI = .020, ΔRMSEA = .006, and ΔSRMR = .008). Ultimately, partial scalar invariance was supported after additionally allowing two item residuals to covary within a country in 13 cases (see Table 7). This final model demonstrated adequate fit to the data and met the change criteria (ΔCFI = .009, ΔRMSEA = .002, and ΔSRMR = .006). As two scalar-invariant items per domain are deemed sufficient to support measurement invariance and compare latent means (Byrne et al., 1989; Steenkamp & Baumgartner, 1998), which was achieved in our case, we proceeded with further analyses.

Applying the same cutoff criteria to the change in model fit between the partially invariant scalar and residuals models, full residual invariance was not supported (ΔCFI = .033, ΔRMSEA = .009, and ΔSRMR = .012). Following the modification indices and allowing 34 more residuals (in total 33%) to freely vary across groups, the model would reach an adequate fit that is not substantially different from the fit of the partially invariant scalar model (ΔCFI = .008, ΔRMSEA = .002, and ΔSRMR = .005). However, within this model, there are cases with more than 50% invariant item

Table 6. Single-group confirmatory factor analysis results.

Country	χ^2	*df*	CFI	TLI	RMSEA	SRMR
Belgium	905.10	341	.92	.90	.07	.06
Germany	739.10	341	.94	.93	.06	.05
Italy	881.42	341	.93	.91	.07	.05
Japan	614.69	341	.95	.94	.06	.04
Korea	619.04	341	.96	.95	.06	.04
Norway	579.20	341	.88	.86	.08	.07
Poland	594.16	341	.93	.92	.05	.05
Portugal	629.92	341	.94	.93	.06	.05
The Netherlands	604.48	341	.93	.91	.06	.06
United States	669.32	341	.94	.93	.06	.05
All	2,526.51	341	.95	.94	.05	.03

Table 7. Multi-group confirmatory factor analysis for test of measurement invariance.

Model	χ^2	*df*	*p*	CFI	RMSEA	SRMR	Invariance
1. Configural	6,840.050	3,410	.000	.931	.054	.052	Yes
2. Metric	7,442.227	3,590	.000	.923	.056	.063	
Δ 1 vs. 2	628.891	180	.000	.008	.002	.011	Yes
3. Scalar	10,575.272	3,770	.000	.864	.073	.082	
Δ 2 vs. 3	3,624.466	180	.000	.059	.017	.019	No
3a. Scalar	8,567.300	3,718	.000	.903	.062	.071	
Δ 2 vs. 3a	1,297.937	128	.000	.020	.006	.008	No
3b. Scalar	8,007.983	3,705	.000	.914	.058	.069	
Δ 2 vs. 3b	652.833	115	.000	.009	.002	.006	Partial
4. Residual	9,891.808	3,914	.000	.881	.067	.081	
Δ 3b vs. 4	1,646.055	209	.000	.033	.009	.012	No
5. Factor mean	9,538.886	3,786	.000	.885	.067	.109	
Δ 3b vs. 5	1,682.011	81	.000	.029	.009	.040	No

Note. N = 3,424.
Model 1: Unconstrained.
Model 2: Constrained all factor loadings.
Model 3: Constrained all factor loadings and all intercepts.
Model 3a: Constrained all factor loadings, constrained intercepts except for 3–7 intercepts per country (52 in total).
Model 3b: Constrained all factor loadings, constrained intercepts except for 3–7 intercepts per country (52 in total), and allowed residuals to covary within countries (WD-1 and WD-2 in Belgium, Germany, Italy, Japan, Norway, Poland, Portugal, and the Netherlands, CC-1 and CC-3 in Belgium, Italy, and the United States, and HM-2 and HM-3 in Germany and Japan).
Model 4: Constrained all factor loadings, constrained intercepts except for 3–7 intercepts per country (52 in total), and allowed residuals to covary within countries (WD-1 and WD-2 in Belgium, Germany, Italy, Japan, Norway, Poland, Portugal, and the Netherlands, CC-1 and CC-3 in Belgium, Italy, and the United States, and HM-2 and HM-3 in Germany and Japan), constrained residuals except those of the items whose intercepts were allowed to vary.
Model 5: Model 3b with the addition of factor means being fixed to zero in all groups.

residuals within at least one domain for eight of the ten countries, indicating that we could not reach partial residual invariance according to the standard threshold (Vandenberg & Lance, 2000). Consequently, we do not consider residual invariance to be achieved. This indicates that the amount of variance in LLWI-SF items not accounted for by the proposed factors is not the same across countries.

Given that partial scalar invariance was established with at least two invariant items for each LLWI-SF subscale, the results indicate that the LLWI-SF model is applicable across diverse contexts and suitable for conducting mean comparisons (Vandenberg & Lance, 2000). Following this, we tested the final model with equal factor means across groups. This model showed a significant decrease in fit relative to the partial scalar invariance model (ΔCFI = .029, ΔRMSEA = .009, and ΔSRMR = .040), indicating that the latent factor means of the LLWI-SF subscales were different across countries. Looking at the pairwise latent mean comparisons, the significant differences for the LLWI-SF domains ranged from 64.4% (individual development) to 77.8% (health management and transition to retirement).

Convergent validity

As shown in Table 5, all LLWI-SF subscales were significantly positively related to age-inclusive HR practices for the overall sample. Correlations ranged between .49 ($p < .001$) and .65 ($p < .001$) and were comparable to those found in the LLWI validation studies (r = [.42, .65] and r = [.43, .73]; Finsel et al., 2023c; Wilckens et al., 2021). On the country level, we also found positive medium to strong correlations between the LLWI-SF subscales and

age-inclusive HR practices (see Supplemental Online Material, https://doi.org/10.48548/pubdata-1716). Consequently, the LLWI-SF demonstrates good convergent validity.

Discriminant validity

Similar to the findings for the LLWI, the LLWI-SF subscales showed a weakly positive significant correlation with positive affect ranging from .12 ($p < .001$) to .27 ($p < .001$) for the overall sample, which was even slightly weaker than the correlations reported for the LLWI ($r = [.15, .31]$ and $r = [.17, .29]$; Finsel et al., 2023c; Wilckens et al., 2021). Likewise, the subscales were only significantly to nonsignificantly weakly correlated with negative affect ranging from −.03 ($p = .07$) to −.17 ($p < .001$), which was also weaker than reported for the LLWI ($r = [-.14, -.29]$ and $r = [-.04, -.24]$; Finsel et al., 2023c; Wilckens et al., 2021). Most of the results at the country level reinforce these findings (see Supplemental Online Material, https://doi.org/10.48548/pubdata-1716). As a result, the LLWI-SF demonstrates good discriminant validity.

Criterion validity

Regarding work engagement, the LLWI-SF subscales showed a weak to moderate positive significant correlation to the criterion for the overall sample, ranging from .19 ($p < .001$) to .42 ($p < .001$). While the correlations were slightly lower than the ones found for the LLWI ($r = [.29, .51]$ and $r = [.32, .47]$; Finsel et al., 2023c; Wilckens et al., 2021), they were still sufficient in size to indicate criterion validity. On the country level, there was also a weak to strong positive correlation between the LLWI-SF subscales and work engagement. However, the correlation was nonsignificant for the *health and retirement coverage* subscale in the Belgian and Norwegian samples. While this finding seems plausible since the domain focuses on retirement benefits and additional insurance and as a result is not as closely linked to the work environment as the other domains, it still diverges from other countries and the LLWI findings.

The correlations between perceived health and the LLWI-SF subscales for the overall sample were significant and positively weak, ranging from .16 ($p < .001$) to .28 ($p < .001$). They were comparable to the findings for the LLWI ($r = [.09, .29]$ and $r = [.21, .32]$; Finsel et al., 2023c; Wilckens et al., 2021). On the country level, we also had weak to moderate positive correlations. However, we found that relationships between perceived health and some LLWI-SF subscales were nonsignificant for the Norwegian, Portuguese, Dutch, and U.S. samples. The only subscales that consistently demonstrated significant correlations were *organizational climate* and *work design*. Surprisingly, the *health management* subscale was not significantly related to perceived health in the Dutch, Norwegian, and U.S. samples.

Regarding post-retirement work intention, with correlations ranging from .22 ($p < .001$) to .34 ($p < .001$), the LLWI-SF subscales demonstrated significant positive relationships with the criterion that were comparable to the validated LLWI subscales ($r = [.11, .41]$ and $r = [.10, .43]$; Finsel et al., 2023c; Wilckens et al., 2021). There was also a positive weak to strong correlation between the LLWI-SF subscales and post-retirement work intention for most individual samples. Only for the Norwegian and Polish samples, there were some nonsignificant relationships with the LLWI-SF subscales *organizational climate*, *work design*, *continued employment*, and *health and retirement coverage*, some of which were also reported in the LLWI validation studies.

Finally, regarding in-role behavior, our findings for the overall sample surprisingly showed slightly weaker correlations for the LLWI-SF subscales ($r = [.01, .19]$) compared to the U.S. validation of the LLWI ($r = [.01, .25]$), even though we used a self-rating instrument instead of the other-rating one as applied by Finsel et al. (2023c). However, contrary to the U.S. LLWI, the relationship was significant for seven of the nine subscales, including the *work design* and *health and retirement coverage* subscales, which did not show a significant correlation for the LLWI. Similar to the LLWI, we could not find a significant correlation between in-role behavior, on the one hand, and the *health management* and *transition to retirement* subscales, on the other hand, for the overall sample. On the country level, we identified severe differences. For example, while all nine LLWI-SF subscales showed a significantly positive correlation to in-role behavior in the Japanese sample, none showed a significant correlation in the Dutch and Norwegian samples. Moreover, significant correlations were positive in all countries except for Portugal. Here, we had two weakly negative correlations with the *transition to retirement* and *health and retirement coverage* subscales.

We further performed regression analyses to test whether the LLWI-SF subscales relate to the investigated criteria above and beyond age-inclusive HR practices. For work engagement, across all countries, all regression models that included the LLWI-SF subscales in addition to the age-inclusive HR practices explained significantly higher amounts of variance compared to models with age-inclusive HR practices only (4.8%–13.7%). For perceived health, the increase in explained variance (0.8%–1.6%) was significant for all countries except Germany, Japan, Norway, and the United States. For the same-employer-post-retirement work intention, there was also an increase in explained variance across all countries (2.4%–13.5%). However, this increase was nonsignificant for Belgium, Italy, Norway, and the Netherlands. Finally, the increase in explained variance for job performance ranged between 0% and 6.8%. It was only significant for Germany, Japan, Poland, and the United States. Moreover, there were some cases where the LLWI-SF subscales were significant negative predictors for the criteria. These cases predominantly referred to the domains *transition to retirement* and *health and retirement coverage* as predictors of post-retirement work intention and job performance.

Overall, the correlation and regression results speak in favor of predictive and, in many cases, also incremental validity of the LLWI-SF for the criteria work engagement, perceived health, and same-employer-post-retirement work intention. Moreover, the findings align with the predictive validity of the LLWI reported by Finsel et al. (2023c). Here, the increase in explained variance compared to a model with age-inclusive HR practices, age-diversity climate (Boehm et al., 2014), and respectful leadership (van Quaquebeke & Eckloff, 2010) was 7% for work engagement, 6% for perceived health, and 4% for same-employer-post-retirement work intention. Regarding job performance, despite significant correlations with the LLWI-SF subscales, we found only limited support for predictive and incremental validity.

Relations with gender, work type, employer size, and industry type

The relationships between the LLWI-SF domains and the demographic and organizational variables for the overall sample are displayed in Table 5. Some significant correlations between the subscales and the gender indicated that for five of the domains (i.e., *work design*, *health management*, *individual development*, *transition to retirement*, and *health and retirement coverage*), men seemed to perceive fewer of these organizational practices in the overall sample (r = [−.04, −.11]). Working in a job with more than 50% office work was weakly negative but significantly correlated to the *organizational climate, work design, health management, knowledge management,* and *continued employment* subscales (r = [−.04, −.08]). On the contrary, working in a job with more than 50% manual work was significantly positively related to the *organizational climate* and *continued employment* subscales (r = .05/.04). Additionally, there was also a weakly negative but significant correlation between primarily manual work and the *leadership*, *work design*, *individual development, transition to retirement,* and *health and retirement coverage* subscales (r = [−.04, −.09]).

Looking at organizational characteristics, the correlations between the LLWI-SF subscales and employer size also depict diverging results. For example, large employer size showed significant positive correlations with the *work design*, *health management*, *individual development*, *transition to retirement,* and *health and retirement coverage* subscales (r = [.04, .10]). This finding, in line with prior research, suggests that some organizational practices are more prevalent in larger organizations. However, only small employer size showed a slight positive correlation of .06 (p < .01) with the *organizational climate* subscale. Regarding the sectors, the service sector was significantly positively related to all LLWI-SF subscales (r = [.04, .11]). On the contrary, the public sector was significantly negatively related to all subscales except *work design* (r = [−.06, −.17]). For the industry sector, there was only a significant positive correlation of .05 (p < .01) with the *health management* subscale.

All in all, even though these relationships were only weak in terms of strength and nonsignificant in many cases, our findings still support the theoretical proposition mentioned by Boehm et al. (2021) as well as by Wilckens et al. (2021) that individual and organizational characteristics play a role for the LLWI-SF domains.

Discussion

This study aimed to develop and validate a short form of the LLWI in ten different countries. We combined qualitative and quantitative strategies to compose the LLWI short form to preserve content validity while also reaching acceptable levels for other psychometric properties of the new measure.

The subscales demonstrated good internal consistency, with the exception of five country-specific LLWI-SF subscales. For the overall sample, Cronbach's alpha ranged from acceptable to very good for all subscales. Even though the values were lower than those of the LLWI, such a decrease in internal consistency is expected considering the subscales' brevity and the elimination of redundant items (Cortina, 1993). In four countries, one or two LLWI-SF subscales had a Cronbach's alpha slightly below the cutoff. Considering that the LLWI-SF is applied in an organizational setting to estimate population effect sizes, the internal consistency is still sufficient (Bonett & Wright, 2015). Moreover, the CFAs revealed the LLWI-SF had a better model fit compared to the LLWI, indicating that the reduced complexity of the model aided in increasing factorial validity. Only in the Norwegian sample did the model not quite meet the requirement for two fit indices. This might be partially caused by the smaller sample size.

Measurement invariance testing revealed that full configural and metric invariance were supported. Moreover, partial scalar invariance was also achieved. In total, ten items had a noninvariant item intercept for at least one of the countries. Roberts et al. (2020) discuss several potential reasons for finding only evidence for partial scalar invariance. One reason the authors mention is that the participants could have understood the items in different ways by attributing different meanings to the words. This could have been the case, for example, for item CE-3 ("*In our organization older applicants are hired as well*") and OC-3 ("*In our organization 'aging' is talked about openly*"), as the perceived meaning of the terms "older" and "aging" could be highly dependent on the cultural context. Age might be understood differently (Kooij et al., 2008), or a different sensitivity could be attributed to age, increasing or decreasing the likelihood of addressing it at work. Another possible reason could be method bias (Roberts et al., 2020), for example, slight differences in the translation of the items or scale labels. Social desirability bias (Edwards, 1957) may have also been a concern, even though the results were not reported to the employers. Indeed, research has shown that social desirability behaviors are related to different cultural dimensions (e.g., Lalwani et al., 2006). Given the number of potentially influencing factors, reaching full scalar measurement invariance is a challenging goal in research, particularly for larger-scale studies (Steinmetz, 2013; Marsh et al., 2018). In response, researchers have proposed that a measure is suitable for cross-cultural research if it has at least two invariant items per scale (Byrne et al., 1989; Steenkamp & Baumgartner, 1998). This was the case for the LLWI-SF subscales in our study, making our measure suitable for cross-cultural research.

The convergent and discriminant validity findings for the LLWI-SF are quite similar to those of the LLWI, although differences exist in individual countries. Overall, except for job performance, the LLWI-SF further demonstrated good criterion validity comparable to the LLWI. On the country level, some findings need further attention in future research, for example, *transition to retirement* was found to be a significant negative predictor of job performance in the Dutch, German, and Portuguese samples. We have also been able to replicate and extend the findings from the LLWI validation studies by demonstrating that the LLWI-SF subscales were related to different individual and organizational characteristics, and in different ways.

Implications for research and practice

There are several benefits to the newly developed LLWI-SF. The main advantage of the LLWI-SF is its reduced administration time. By shortening the 80-item measure to 29 items, we shortened its length by 64%, reducing the administration time by more than half. Comparing its overall sample reliability to Wilckens et al. (2021) and Finsel et al. (2023c), the average loss in internal consistency per domain ranged between 2% and 17%. This constitutes a reasonable trade-off between gained efficiency and reduced reliability.

Even though we had to sacrifice the indicator-level factor structure, we ensured that the LLWI-SF still covers the indicator constructs by keeping at least one item per indicator. This approach, which is based on a qualitative strategy and emphasizes preserving content validity, has been shown to be preferable to strategies that are purely quantitative (e.g., Kruyen et al., 2013; Stanton et al., 2002).

Potential fields of application for the LLWI-SF are broad. For research, the reduced administration time of the 29-item measure opens up new opportunities to incorporate a comprehensive measure of organizational practices for older workers in surveys investigating the phenomenon of aging at the workplace and examining its determinants and outcomes across occupational sectors and around the globe. This allows for more longitudinal research, which is needed to adequately study the impact of organizational practices on older workers' attitudes and behaviors over time (Boehm et al., 2021). Through the development and validation of the LLWI-SF, we facilitate the collection and accumulation of such longitudinal data to study the transition to more age-friendly organizations. Furthermore, the LLWI-SF still allows the use of individual subscales, offering scales with only three to four items to measure specific practices.

As the LLWI-SF demonstrates partial scalar measurement invariance across the ten countries, the potential for cross-country comparisons is a major advantage of the LLWI-SF for future research. Indeed, our findings indicate that there are latent factor mean differences across countries. However, there has so far been little comparative empirical research focused on the meso-macro level relationship between organizational practices for older workers and country-level factors (Finsel et al., 2023b). In this regard, scholars can use the LLWI-SF to investigate and understand potential differences in the availability of these organizational practices and their implications for older workers and organizations across countries.

Practitioners can use the LLWI-SF for efficient prescreening to identify areas for in-depth assessment and organizational interventions (e.g., using the LLWI). The LLWI-SF might be particularly valuable for organizations that have only started to deal with and engage with their aging workforce. Moreover, as labor shortages continue to rise (Farr-Wharton et al., 2023), monitoring changes in organizations and evaluating the effectiveness of organizational interventions to create a more age-friendly workplace will become increasingly important. As the LLWI is constrained in its use due to its significant time and financial costs, the LLWI-SF allows organizations to examine changes over time by collecting reliable data quickly and easily. The LLWI-SF also offers the possibility for benchmarking, enabling organizations to compare themselves internally among different sites or externally among peers in the broader industry. Policy-makers can also use the shortened LLWI-SF more efficiently to aid in their decision-making processes and strategy development.

Limitations and future research directions

It is important to note that our study is not without limitations. First, even though our samples show good coverage of demographic and organizational characteristics, data were collected through online panels and survey platforms and were not fully population-representative. However, several studies have shown that results from such samples are comparable to those from conventional sampling methods such as organizational samples (e.g., Walter et al., 2019). Moving forward, more representative sampling techniques should be applied, particularly in cross-cultural comparative research. Moreover, some of our findings on the country level diverge from the general trend (e.g., comparably low reliabilities for specific subscales or the model fit in Norway) and warrant further investigation. Replication studies are needed to strengthen the evidence for the reliability and validity of the LLWI-SF in some of the countries investigated.

Given the variation in cross-cultural differences, even though we collected data in ten countries, validating the LLWI-SF in further countries could shed new light on the construct. For example, including countries with low retirement ages like Turkey (OECD, 2023) could test the robustness of the measure in diverging contexts and open up new opportunities for cross-cultural research. In light of country-specific age structures, retirement ages, and legislation, it would be of particular interest to test the cultural adequateness of defining a universal target group of the LLWI-SF as one comprising employees aged 50 and older. For example, a lower age restriction might be more appropriate in countries with a generally younger working population, like Turkey or Nigeria (Aksaray & Marcus, 2025; United Nations, 2022).

Since the aim of our study was a short-scale validation, we used cross-sectional data to test for convergent, discriminant, and criterion validity. We encourage future research wherein the predictive validity of the LLWI domains in the light of perceptual, attitudinal, and behavioral employee outcomes, on the one hand, and organizational outcomes, on the other hand, are investigated. Regarding employee outcomes, the sustainable career framework (de Vos et al., 2020), with its three domains of health, happiness, and productivity (van der Heijden, 2005), might be a promising avenue. Sickness absence rates, average retirement ages, profit, and employee turnover rates are examples of interesting organizational outcomes. Moreover, necessary condition analysis (Dul et al., 2023) could be used to determine which LLWI domains are "conditio sine qua non" for specific outcomes.

The brevity of the LLWI-SF can also facilitate more extensive research projects and surveys. This includes comparative studies that investigate patterns of relationships between different categories of workers (e.g., age, gender, or occupational type) across countries and industry sectors. In particular, future research could delve deeper into incorporating differences and similarities between countries with regard to contextual factors (e.g., normal retirement ages, labor force participation rates; OECD, 2023) to improve the understanding of potential differences in the prevalence of organizational practices for older workers. Studies have shown, for example, that legislation can influence employers' decisions to implement such practices (e.g., Ollier-Malaterre et al., 2013). Investigating contextual factors at both the predictor and outcome level of organizational practices for older workers could help to better understand the emergence and implications of such differences (Boehm et al., 2021; Finsel et al., 2023b). Finally, in order to screen organizational practices relevant to older workers in a more population-based research context, we recommend the development of an LLWI version that captures each of the nine domains with a single item.

Conclusion

The LLWI-SF is a measurement instrument that is reliable and valid across different countries. With this short form,

we provide practitioners and researchers with a multidimensional and efficient measure available in ten languages to assess practices relevant to organizations with an increasingly aging workforce.

Data availability

The data underlying this article will be shared on reasonable request to the corresponding author.

Author contributions

Julia Finsel (Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Supervision, Visualization, Writing—original draft, Writing—review & editing), Hila Axelrad (Conceptualization, Methodology, Writing—original draft, Writing—review & editing), Su Jung Choi (Conceptualization, Investigation, Methodology, Writing—original draft, Writing—review & editing), Eva Derous (Conceptualization, Investigation, Methodology, Writing—original draft, Writing—review & editing), Xiuzhu Gu (Conceptualization, Funding acquisition, Investigation, Methodology, Writing—original draft, Writing—review & editing), Paolo Guandalini (Conceptualization, Funding acquisition, Investigation, Methodology, Writing—original draft, Writing—review & editing), Jaeyoung Ha (Conceptualization, Investigation, Methodology, Writing—original draft, Writing—review & editing), Eun Seok Kim (Investigation, Writing—original draft, Writing—review & editing), Izabela Marzec (Conceptualization, Investigation, Methodology, Writing—original draft, Writing—review & editing), Reidar Mykletun (Conceptualization, Funding acquisition, Investigation, Methodology, Writing—original draft, Writing—review & editing), Eduardo Oliveira (Conceptualization, Funding acquisition, Investigation, Methodology, Writing—original draft, Writing—review & editing), Sofija Pajic (Conceptualization, Formal analysis, Investigation, Methodology, Writing—original draft, Writing—review & editing), Maaike Schellaert (Conceptualization, Investigation, Methodology, Writing—original draft, Writing—review & editing), Beatrice Van der Heijden (Conceptualization, Investigation, Methodology, Writing—original draft, Writing—review & editing), Michela Vignoli (Conceptualization, Investigation, Methodology, Writing—original draft, Writing—review & editing), Anne M. Wöhrmann (Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Visualization, Writing—original draft, Writing—review & editing), and Jürgen Deller (Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing—original draft, Writing—review & editing).

Funding

This research was partially supported by the European Union - Next Generation EU (grant ID E63C22001230001), the ias-Gruppe, the Fundação para a Ciência e a Tecnologia (UIDB/04105/2020), the Japan Society for the Promotion of Science (grant number 18H01654), and the Norwegian Centre for Senior Policies, Akersgata 32, 0180 Oslo.

Conflicts of interest: None declared.

Appendix

LLWI-SF

Table A1. LLWI-SF items.

Code	English-language item
OC organizational climate	
OC-1	In our organization, regardless of age, all employees have the same opportunities.
OC-2	In our organization, there is a positive attitude toward older employees.
OC-3	In our organization, "aging" is talked about openly.
LE leadership	
LE-1	Managers of our organization show appreciation both for current work results as well as for the overall performance of their employees.
LE-2	Managers of our organization sincerely support their employees in their professional and personal development.
LE-3	Managers of our organization are interested in the well-being of their employees.
WD work design	
WD-1	Employees of our organization have enough flexibility in their working time organization to appropriately address their personal needs.
WD-2	Employees of our organization can choose their place of work to ensure a good balance between their work and private life (work–life balance).
WD-3	In our organization, managers change the tasks of their employees in the foreseeable future (e.g., within half a year) if the tasks no longer correspond to the employee's ability to perform and to withstand stress.
WD-4	In our organization, workplaces are designed according to ergonomic recommendations.
HM health management	
HM-1	Employees of our organization are encouraged to move as much as possible in the workplace (e.g., use the stairs, take a walk during lunch break, exercise during lunch break, use the bicycle to work).
HM-2	In our organization, employees receive therapeutic help in the workplace or in the immediate vicinity if required (e.g., physiotherapy in case of great physical stress and strain, stress management trainings).

Code	English-language item
HM-3	In our organization, employees are made aware of health-promoting behavior (e.g., through training, counseling, displays).
ID individual development	
ID-1	In our organization, individual development prospects and qualification requirements are identified with employees, regardless of age.
ID-2	In our organization, older employees are offered training to learn new competencies and develop their expertise.
ID-3	In our organization, employees move to a different job or position if it better suits their specific skills and abilities.
KM knowledge management	
KM-1	In our organization, there are processes/procedures to systematically pass on the knowledge and experience of older employees to their younger colleagues before they leave the organization.
KM-2	In our organization, there are regular opportunities for every employee to exchange experiences and knowledge (e.g., in regular meetings).
KM-3	In our organization, older and younger employees are encouraged to share their knowledge and experience.
TR transition to retirement	
TR-1	In our organization, managers discuss early with their employees (e.g., from the age of 55) as to how to make the transition to retirement.
TR-2	In our organization, the transition to retirement is flexibly shaped according to employee needs.
TR-3	Our organization offers counseling to employees who are about to retire so they can reflect upon their expectations and plans for retirement.
TR-4	Our organization maintains active contact with retired employees (e.g., by an alumni network).
CE continued employment	
CE-1	In our organization, employment opportunities for people in retirement age are clearly defined and structured (e.g., by integration into strategic workforce planning).
CE-2	In our organization, working conditions (time and type of activity) for employees in retirement age are flexibly adapted to their wishes.
CE-3	In our organization, older applicants are hired as well.
RC health and retirement coverage	
RC-1	Our organization offers employees comprehensive opportunities to save money for their retirement (e.g., retirement plans offered by employer).
RC-2	Our organization offers employees good personal advice on financial security in later life.
RC-3	Our organization offers employees private supplemental insurance as part of the total remuneration package (e.g., additions to health or long-term care insurance, occupational disability).

Note. Published with kind permission of Max R. Wilckens, Anne M. Wöhrmann, Julia S. Finsel, and Jürgen Deller 2023. All Rights Reserved.

References

Adams, G. A., & Beehr, T. A. (1998). Turnover and retirement: A comparison of their similarities and differences. *Personnel Psychology 51*(3), 643–665. https://doi.org/10.1111/j.1744-6570.1998.tb00255.x

Aksaray, G., & Marcus, J. (2025). Bridge employment in middle vs. late life for men and women: Gendered social roles or life structures? *Work, Aging and Retirement*. Advance online publication. https://doi.org/10.1093/workar/waae019

Axelrad, H., Kalev, A., & Lewin-Epstein, N. (2024). Ambivalent bias at work: Managers' perceptions of older workers across organizational contexts. *Work, Employment and Society 38*(4), 1148–1171. https://doi.org/10.1177/09500170231175790

Aycan, Z., Kanungo, R., Mendonca, M., Yu, K., Deller, J., Stahl, G., & Kurshid, A. (2000). Impact of culture on human resource management practices: A 10-country comparison. *Applied Psychology 49*(1), 192–221. https://doi.org/10.1111/1464-0597.00010

Aycan, Z., Kanungo, R. N., & Sinha, J. B. P. (1999). Organizational culture and human resource management practices: The model of culture fit. *Journal of Cross-Cultural Psychology 30*(4), 501–526. https://doi.org/10.1177/0022022199030004006

Boehm, S. A., Kunze, F., & Bruch, H. (2014). Spotlight on age-diversity climate: The impact of age-inclusive HR practices on firm-level outcomes. *Personnel Psychology 67*(3), 667–704. https://doi.org/10.1111/peps.12047

Boehm, S. A., Schröder, H., & Bal, M. (2021). Age-related human resource management policies and practices: Antecedents, outcomes, and conceptualizations. *Work, Aging and Retirement 7*(4), 257–272. https://doi.org/10.1093/workar/waab024

Bonett, D. G., & Wright, T. A. (2015). Cronbach's alpha reliability: Interval estimation, hypothesis testing, and sample size planning. *Journal of Organizational Behavior 36*(1), 3–15. https://doi.org/10.1002/job.1960

Briggs, S. R., & Cheek, J. M. (1986). The role of factor analysis in the development and evaluation of personality scales. *Journal of Personality 54*(1), 106–148. https://doi.org/10.1111/j.1467-6494.1986.tb00391.x

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology 1*(3), 185–216. https://doi.org/10.1177/135910457000100301

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

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for 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

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

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates, Publishers.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology 78*(1), 98–104. https://doi.org/10.1037//0021-9010.78.1.98

Curran, P. J., West, S. G., & Finch, J. F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods 1*(1), 16–29. https://doi.org/10.1037//1082-989x.1.1.16

de Vos, A., van der Heijden, B. I. J. M., & Akkermans, J. (2020). Sustainable careers: Towards a conceptual model. *Journal of Vocational Behavior*, *117*, 103196. https://doi.org/10.1016/j.jvb.2018.06.011.

de Wind, A., Geuskens, G. A., Ybema, J. F., Blatter, B. M., Burdorf, A., Bongers, P. M., & van der Beek, A. J. (2014). Health, job characteristics, skills, and social and financial factors in relation to early retirement – Results from a longitudinal study in the Netherlands. *Scandinavian Journal of Work, Environment and Health 40*(2), 186–194. https://doi.org/10.5271/sjweh.3393

Dul, J., Hauff, S., & Bouncken, R. B. (2023). Necessary condition analysis (NCA): Review of research topics and guidelines for good practice. *Review of Managerial Science 17*(2), 1535–1537. https://doi.org/10.1007/s11846-023-00634-z

Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Dryden Press.

Eisele, G., Vachon, H., Lafit, G., Kuppens, P., Houben, M., Myin-Germeys, I., & Viechtbauer, W. (2022). The effects of sampling frequency and questionnaire length on perceived burden, compliance, and careless responding in experience sampling data in a student population. *Assessment 29*(2), 136–151. https://doi.org/10.1177/1073191120957102

Eisenberger, R., Armeli, S., Rexwinkel, B., Lynch, P. D., & Rhoades, L. (2001). Reciprocation of perceived organizational support. *The Journal of Applied Psychology 86*(1), 42–51. https://doi.org/10.1037/0021-9010.86.1.42

Farr-Wharton, B., Bentley, T., Onnis, L.-a., Caponecchia, C., Neto, A. D. A., O'Neill, S., & Andrew, C. (2023). Older worker-orientated human resource practices, wellbeing and leave intentions: a conservation of resources approach for ageing workforces. *International Journal of Environmental Research and Public Health 20*(3), 2725. https://doi.org/10.3390/ijerph20032725.

Finsel, J. S., Venz, L., Wöhrmann, A. M., Wilckens, M. R., & Deller, J. (2023a). Worlds apart: does perceptual congruence between leaders and older employees regarding age-friendly organizational climate, management, and work design matter? *Work, Aging and Retirement 10*(2), 123–137. https://doi.org/10.1093/workar/waad009

Finsel, J. S., Wöhrmann, A. M., & Deller, J. (2023b). A conceptual cross-disciplinary model of organizational practices for older workers: Multilevel antecedents and outcomes. *The International Journal of Human Resource Management 34*(22), 4344–4396. https://doi.org/10.1080/09585192.2023.2199939

Finsel, J. S., Wöhrmann, A. M., Wang, M., Wilckens, M. R., & Deller, J. (2023c). Later life workplace index: validation of an English version. *Work, Aging and Retirement 9*(1), 71–94. https://doi.org/10.1093/workar/waab029

Fisher, G. G., Chaffee, D. S., & Sonnega, A. (2016). Retirement timing: A review and recommendations for future research. *Work, Aging and Retirement 2*(2), 230–261. https://doi.org/10.1093/workar/waw001

Froidevaux, A., Jovanovska, I., & Masdonati, J. (2024). Maintenir les employés séniors en emploi jusqu'à l'âge de la retraite et au-delà [Maintaining older workers at work until retirement age and beyond]. In C. Helson, E. Loarer, & J.-L. Bernaud (Eds.), *Psychologie de la Retraite [Psychology of Retirement]* (pp. 95–110). Dunod. https://www.dunod.com/sciences-humaines-et-sociales/psychologie-retraite

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly 73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Hallpike, H., Van der Heijden, B. I. J. M., & Vallée-Tourangeau, G. (2025). What do you do when your career script runs out? How older workers decide whether and how to sustain their careers. *Work, Aging and Retirement 11*(1), 97–118. https://doi.org/10.1093/workar/waad027

Hertel, G., & Zacher, H. (2018). Managing the aging workforce. In D. S. Ones, N. Anderson, H. K. Sinangil, & C. Viswesvaran (Eds.), *The SAGE handbook of industrial, work & organizational psychology* (pp. 396–428). Sage.

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

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods 14*(1), 6–23. https://doi.org/10.1037/a0014694

Jonsson, R., Lindegård, A., Björk, L., & Nilsson, K. (2020). Organizational hindrances to the retention of older healthcare workers. *Nordic Journal of Working Life Studies 10*(1), 41–58. https://doi.org/10.18291/njwls.v10i1.118679

Kline, R. B. (2005). *Principles and practice of structure equation modeling* (2nd ed.). Guildford Press.

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine 15*(2), 155–163. https://doi.org/10.1016/j.jcm.2016.02.012

Kooij, D., de Lange, A., Jansen, P., & Dikkers, J. (2008). Older workers' motivation to continue to work: Five meanings of age. *Journal of Managerial Psychology 23*(4), 364–394. https://doi.org/10.1108/02683940810869015

Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing 13*(3), 223–248. https://doi.org/10.1080/15305058.2012.703734

Laguerre, R. A., & Barnes-Farrell, J. L. (2024). Bringing self-determination theory to the forefront: Examining how human resource practices motivate employees of all ages to succeed. *Journal of Business and Psychology 40*(1), 1–37. https://doi.org/10.1007/s10869-024-09951-w

Lalwani, A. K., Shavitt, S., & Johnson, T. (2006). What is the relation between cultural orientation and socially desirable responding? *Journal of Personality and Social Psychology 90*(1), 165–178. https://doi.org/10.1037/0022-3514.90.1.165

Liden, R. C., Wayne, S. J., Meuser, J. D., Hu, J., Wu, J., & Liao, C. (2015). Servant leadership: Validation of a short form of the SL-28. *The Leadership Quarterly 26*(2), 254–269. https://doi.org/10.1016/j.leaqua.2014.12.002

Marcus, J., Scheibe, S., Kooij, D., Truxillo, D. M., Zaniboni, S., Abuladze, L., al Mursi, N., Bamberger, P. A., Balytska, M., Betanzos, N. D., asJ.Perek-Białas, J., Boehm, S. A., Burmeister, A., Cabib, I., Caon, M., Deller, J., Derous, E., Drury, L., Eppler-Hattab, R., … Žnidaršič, J. (2024). LeverAge: A European network to leverage the multi-age workforce. *Work, Aging and Retirement 10*(4), 309–316. https://doi.org/10.1093/workar/waae009

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

Matheis, T. K., de Campos, S. A. P., Vieira, K. M., Lehnhart, E., dos, R., & Estivalete, V. F. B. (2024). Organizational climate scale for

public service: Development and validation. *Administrative Sciences 14*(5). https://doi.org/10.3390/admsci14050090. Article 90

McDonald, R. P., & Ho, M. -H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods 7*(1), 64–82. https://doi.org/10.1037/1082-989X.7.1.64

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *The Journal of Applied Psychology 93*(3), 568–592. https://doi.org/10.1037/0021-9010.93.3.568

Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods 17*(4), 372–411. https://doi.org/10.1177/1094428114548590

Newman, D. A., & Cottrell, J. M. (2015). Missing data bias: Exactly how bad is pairwise deletion? In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends: Doctrine, verity and fable in organizational and social sciences* (pp. 133–161). Routledge.

Nishii, L. H., Khattab, J., Shemla, M., & Paluch, R. M. (2018). A multilevel process model for understanding diversity practice effectiveness. *Academy of Management Annals 12*(1), 37–82. https://doi.org/10.5465/annals.2016.0044

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

OECD. (2023). *Pensions at a glance 2023: OECD and G20 indicators*. OECD Publishing. https://doi.org/10.1787/678055dd-en

Ollier-Malaterre, A., McNamara, T., Matz-Costa, C., Pitt-Catsouphes, M., & Valcour, M. (2013). Looking up to regulations, out at peers or down at the bottom line: How institutional logics affect the prevalence of age-related HR practices. *Human Relations 66*(10), 1373–1395. https://doi.org/10.1177/0018726713478244

Pilipiec, P., Groot, W., & Pavlova, M. (2021). The effect of an increase of the retirement age on the health, well-being, and labor force participation of older workers: A systematic literature review. *Journal of Population Ageing 14*(2), 271–315. https://doi.org/10.1007/s12062-020-09280-9

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: DR 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Rammstedt, B., & Beierlein, C. (2014). Can't we make it any shorter? The limits of personality assessment and way to overcome them. *Journal of Individual Differences 35*(4), 212–220. https://doi.org/10.1027/1614-0001/a000141

Reich, A. J., Claunch, K. D., Verdeja, M. A., Dungan, M. T., Anderson, S., Clayton, C. K., & Thacker, E. L. (2020). What does "successful aging" mean to you? — Systematic review and cross-cultural comparison of lay perspectives of older adults in 13 countries, 2010–2020. *Journal of Cross-Cultural Gerontology 35*(4), 455–478. https://doi.org/10.1007/s10823-020-09416-6

Revelle, W. (2024). *Using R and the psych package to find ω*. Department of Psychology, Northwestern University. http://personality-project.org/r/psych/HowTo/omega.pdf

Riekhoff, A. -J. (2024). Employers' retirement age norms in European comparison. *Work, Aging and Retirement 10*(4), 317–330. https://doi.org/10.1093/workar/waad015

Roberts, C., Sarrasin, O., & Ernst Stähli, M. (2020). Investigating the relative impact of different sources of measurement non-equivalence in comparative surveys: An illustration with scale format, data collection mode and cross-national variations. *Survey Research Methods 14*(4), 399–415. https://doi.org/10.18148/srm/2020.v14i4.7416

Schaufeli, W. B., Shimazu, A., Hakanen, J., Salanova, M., & de Witte, H. (2019). An ultra-short measure for work engagement. The UWES-3 validation across five countries. *European Journal of Psychological Assessment 35*(4), 577–591. https://doi.org/10.1027/1015-5759/a000430

Simmering, M. J., Fuller, C. M., Richardson, H. A., Ocal, Y., & Atinc, G. M. (2015). Marker variable choice, reporting, and interpretation in the detection of common method variance. *Organizational Research Methods 18*(3), 473–511. https://doi.org/10.1177/1094428114560023

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment 12*(1), 102–111. https://doi.org/10.1037//1040-3590.12.1.102

Stanton, J. M., Sinar, E. F., Balzer, W. K., & Smith, P. C. (2002). Issues and strategies for reducing the length of self-report scales. *Personnel Psychology 55*(1), 167–194. https://doi.org/10.1111/j.1744-6570.2002.tb00108.x

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

Steinmetz, H. (2013). Analyzing observed composite differences across groups: Is partial measurement invariance enough? *Methodology 9*(1), 1–12. https://doi.org/10.1027/1614-2241/a000049

Stoyanova, S., Giannouli, V., Miteva, S., & Ivantchev, N. (2024). How different forms of knowledge transfer at workplace are related to self-rated work performance? *African Journal of Biomedical Research 27*(4S), 2434–2453. https://doi.org/10.53555/AJBR.v27i4S.3312

Taneva, S. K., & Arnold, J. (2018). Thriving, surviving and performing in late career: A mixed-method study of pathways to successful aging in organizations. *Work, Aging and Retirement 4*(2), 189–212. https://doi.org/10.1093/workar/wax027

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the Positive and Negative Affect Schedule (PANAS). *Journal of Cross-Cultural Psychology 38*(2), 227–242. https://doi.org/10.1177/0022022106297301

United Nations. (2022). *World population prospects 2022: Summary of results*. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/wpp2022_summary_of_results.pdf

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: an overview. *European Review of Applied Psychology 54*(2), 119–135. http://dx.doi.org/10.1016/j.erap.2003.12.004

van der Heijden, B. I. J. M. (2005). "*No one has ever promised you a rose garden.*" *On shared responsibility and employability enhancing strategies throughout careers*. Van Gorcum.

van der Heijden, B. I. J. M., Notelaers, G., Peters, P., Stoffers, J. M. M., de Lange, A. H., Froehlich, D. E., & van der Heijde, C. M. (2018). Development and validation of the short-form employability five-factor instrument. *Journal of Vocational Behavior 106*(2017), 236–248. https://doi.org/10.1016/j.jvb.2018.02.003

van Quaquebeke, N., & Eckloff, T. (2010). Defining respectful leadership: What it is, how it can be measured, and another glimpse at what it is related to. *Journal of Business Ethics 91*(3), 343–358. https://doi.org/10.1007/s10551-009-0087-z

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

Walter, S. L., Seibert, S. E., Goering, D., & O'Boyle, E. H. (2019). A tale of two sample sources: do results from online panel data and conventional data converge? *Journal of Business and Psychology 34*(4), 425–452. https://doi.org/10.1007/s10869-018-9552-y

Wilckens, M. R., Wöhrmann, A. M., Adams, C., Deller, J., & Finkelstein, R. (2020). Integrating the German and the US perspective on organizational practices for later life work: the later life work index. In S. J. Czaja, J. Sharit, & J. B. James (Eds.), *Current and emerging trends in aging and work* (pp. 59–79). Springer.

Wilckens, M. R., Wöhrmann, A. M., Deller, J., & Finsel, J. (2023). Health and the intention to retire: Exploring the moderating effects of human resources practices. *The International Journal of Human Resource Management 34*(18), 3520–3554. https://doi.org/10.1080/09585192.2022.2133967

Wilckens, M. R., Wöhrmann, A. M., Deller, J., & Wang, M. (2021). Organizational practices for the aging workforce: Development

and validation of the later life workplace index. *Work, Aging and Retirement* 7(4), 352–386. https://doi.org/10.1093/workar/waaa012

Wöhrmann, A. M., Deller, J., & Pundt, L. (2018). Complementing AAI at the meso level: The silver work index. In A. Zaidi, S. Harper, K. Howse, G. Lamura, & J. Perek-Bialas (Eds.), *Building evidence for active ageing policies. Active Ageing Index and its potential* (pp. 75–94). Palgrave Macmillan.

Wöhrmann, A. M., Deller, J., & Wang, M. (2013). Outcome expectations and work design characteristics in post-retirement work planning. *Journal of Vocational Behavior 83*(3), 219–228. https://doi.org/10.1016/j.jvb.2013.05.003

Yuan, K.-H., & Bentler, P. M. (2007). Robust procedures in structural equation modeling. In S.-Y. Lee (Ed.), *Handbook of computing and statistics with applications* (Vol. *1*, pp. 367–397). Elsevier. https://doi.org/10.1016/B978-044452044-9/50020-3

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales – Five misunderstandings and ways to overcome them. *Journal of Individual Differences 35*(4), 185–189. https://doi.org/10.1027/1614-0001/a000148