



# Psychometric Analysis of a German-Language Version of the Work–Family Conflict and Family–Work Conflict Scale

Nikola Komlenac\*†, Lisa Stockinger, Tanja Vogler† and Margarethe Hochleitner

Gender Medicine Unit, Medical University of Innsbruck, Innsbruck, Austria

## OPEN ACCESS

### Edited by:

João P. Marôco,  
University Institute of Psychological,  
Social and Life Sciences (ISPA),  
Portugal

### Reviewed by:

Sónia P. Gonçalves,  
University of Lisbon, Portugal  
Yu-Yu Hsiao,  
University of New Mexico,  
United States

### \*Correspondence:

Nikola Komlenac  
nikola.komlenac@i-med.ac.at

### †ORCID:

Nikola Komlenac  
orcid.org/0000-0002-2647-4461  
Tanja Vogler  
orcid.org/0000-0003-2261-898X

### Specialty section:

This article was submitted to  
Quantitative Psychology  
and Measurement,  
a section of the journal  
Frontiers in Psychology

**Received:** 24 September 2021

**Accepted:** 08 December 2021

**Published:** 23 December 2021

### Citation:

Komlenac N, Stockinger L,  
Vogler T and Hochleitner M (2021)  
Psychometric Analysis of a  
German-Language Version of the  
Work–Family Conflict  
and Family–Work Conflict Scale.  
*Front. Psychol.* 12:782618.  
doi: 10.3389/fpsyg.2021.782618

The Work–Family Conflict and Family–Work Conflict Scale (WFC & FWC Scale) is a questionnaire commonly used to assess conflicts that arise when required time devotion and strain for work obligations interfere with family responsibilities (work–family conflict) and conflicts that arise when family responsibilities interfere with work responsibilities (family work conflict). Past reports on the psychometric properties and recommendations for application of the WFC & FWC Scale mostly rely on samples from the United States. The current study is the first to report psychometric properties of a German-language version of the WFC & FWC Scale, including invariance analyses across women and men, and test-retest reliabilities. The analysis of the latent structure that was based on responses from 274 employees (77.0% women, 23.0% men) of a medical university in Austria revealed that the bifactor model had a satisfactory fit with the data. Configural and metric invariance indicated a similar factor structure and similar meaning in women and men. However, scalar invariance cannot be assumed. Thus, differences in scale scores between women and men might not adequately reflect level differences in the underlying latent factor. High internal consistencies and high test-retest reliabilities offer evidence for adequate reliability. Additionally, evidence for convergent (links to work stress and relationship satisfaction) and divergent validity (no links to career ambition) were found. In summary, the current study offers adequate evidence for validity and reliability of a German-language version of the WFC & FWC Scale.

**Keywords:** Work–Family Conflict and Family–Work Conflict Scale, psychometric analysis, invariance analysis, confirmatory factor analysis (CFA), German translation, validity, reliability

## INTRODUCTION

The Work–Family Conflict and Family–Work Conflict Scale (WFC & FWC Scale) (Netemeyer et al., 1996) is an instrument that is widely used to assess the so-called work-family conflict (WFC) and family work conflict (FWC) (Min et al., 2021). WFC arises when general demands of time devoted to and strain created by work interfere with family responsibilities (Greenhaus and Beutell, 1985; Netemeyer et al., 1996). FWC can be caused by family responsibilities setting time demands or causing strain that interfere with work responsibilities (Greenhaus and Beutell, 1985; Netemeyer et al., 1996). The experience of strong WFC and/or strong FWC has been associated with negative work-related outcomes (such as lower job performance, reduced job satisfaction, or feelings of

exhaustion), family related outcomes (e.g., reduced leisure satisfaction, more frequent conflicts with family members), or stress/health-related outcomes (e.g., heightened anxiety, depression, and poorer general health) (Amstad et al., 2011; Sirgy and Lee, 2018; Borgmann et al., 2019; Borgmann et al., 2020).

The WFC & FWC Scale is a brief instrument that focuses mostly on assessing conflicting time demands and experienced strain caused by conflicting demands of work and family responsibilities (Netemeyer et al., 1996). Confirmatory factor analyses based on data from three different United States samples show that two-factor models had satisfactory fit indices (comparative fit index (CFI) >0.93; Tucker–Lewis Index (TLI) >0.91). Furthermore, the scales had acceptable internal consistencies (Cronbach's alphas >0.83) (Netemeyer et al., 1996). The advantage of the WFC & FWC Scale is its short length, while still measuring two constructs, namely WFC and FWC. Furthermore, in comparison to other instruments that measure WFC (Carlson et al., 2000; Grzywacz et al., 2006), the WFC & FWC Scale has proven to have satisfactory measurement precision, especially in heterogeneous samples and when participants experience moderate levels of WFC or FWC (Min et al., 2021).

However, to authors' knowledge psychometric analyses of a German-language version of the WFC & FWC Scale have not been reported. To close this gap in the literature, the current study reports evidence for validity and reliability of a German-language version of the WFC & FWC Scale. To this end, the WFC & FWC Scale (Netemeyer et al., 1996) was translated from English to German by a German native-speaker and professional translator. In order to test the face validity of this translation, an English native-speaker and professional translator translated the German version back to English (back-translation method) and the two English-language versions were compared (Brislin, 1970). Authors discussed any discrepancies until they reached consensus on the final wording. All the other questionnaires that were used in the current study to estimate convergent and divergent validity were developed in English (Cavanaugh et al., 2000; Funk and Rogge, 2007; Meeussen and Van Laar, 2018) and were also translated to German with the forth-and-back procedure (Brislin, 1970).

Specific indicators of reliability and validity (Döring and Bortz, 2016) that were previously estimated for the original English version of the WFC & FWC Scale have been estimated for the German-language version of the WFC & FWC Scale. First, the internal factor structure of the instrument was analyzed and internal consistencies were calculated (Netemeyer et al., 1996). Working from known associations, convergent validity was tested by calculating the correlation between the WFC & FWC Scale and experienced stress at work or relationship satisfaction. Divergent validity was tested by analyzing correlations between the WFC & FWC Scale and career ambitions (Netemeyer et al., 1996; Ellinas et al., 2018). The current study adds further to the current literature by providing test-retest reliability and by performing invariance analyses across women and men. Analyzed was whether the German-language version of the WFC & FWC Scale measures similar constructs, with similar meaning, similar levels of the latent factors, and a similar degree of precision

in women and men (Kline, 2015). Knowing whether invariance can be assumed across women and men is important for the interpretation of previously reported gender differences on the WFC & FWC Scale (Shockley et al., 2017).

## METHODS

### Procedures and Participants

This study was part of a larger study that was conducted from April to June 2020 (T1) among employees at an Austrian medical university (Komlenac et al., 2021). In order to estimate test-retest reliability, data were collected at a second time point from September to October 2020 (T2). Data sets were matched with the help of subject-generated identification codes (Schnell et al., 2010). All employees who followed the e-mailed invitation were asked to give informed consent before opening the online questionnaire that was hosted on *SoSci: der onlineFragebogen*<sup>1</sup>. The medical university's Ethics Committee exempted the current study from full ethics review and confirmed (on April 8, 2020) that under Austrian law the current study did not require formal approval by an ethics committee [Federal Act on the Organization of Universities and their Studies (BGBl, 2002)], Hospitals and Health Resorts Act [Bundesgesetz über Krankenanstalten und Kuranstalten (KAKuG), 2016].

In total, 400 employees participated in the study at T1. Of those respondents, 274 were included for the analyses of the internal factor structure, the invariance analyses, and the convergent and divergent validity, because they passed two attention-check items ("Please select the response "Agree") (Huang et al., 2012) and did not discontinue the survey before reaching the WFC & FWC Scale. The current study's sample size seemed adequate, because lower bound sample size calculations revealed that a minimum sample size of  $n \geq 156$  is recommended for detecting effects of at least 0.4 with a power of 0.8 at a significance level of 0.05 in structural equation models with three latent variables and ten indicator variables (Westland, 2010). Furthermore, the current study's sample size complies with Barrett (2007) recommendation that structural equation models should have a sample size of at least 200. Such a sample size corresponds with the median sample sizes previously used for calculation of structural equation models (MacCallum and Austin, 2000; Shah and Goldstein, 2006). The final sample included 211 women (77.0%) and 63 men (23.0%). Participants were on average 41.3 ( $SD = 10.4$ ) years old. The majority of the participants held a university degree (71.4%), worked full time (64.2%), and in academia (54.0%). The other participants had absolved secondary school, vocational training or had a university entrance-level qualification (28.6%), worked part-time (35.8%), and in administration (46.0%). Most of the participants identified as heterosexual (93.0%), had Austrian nationality (79.6%), were in a relationship (85.4%), and had children (55.8%). The other participants did not identify as heterosexual (7.0%), were not Austrian (20.4%), were single (14.6%), and did not

<sup>1</sup><http://soscisurvey.de/>

have children (44.2%). Responses from those 82 participants (17.1% men, 82.9% women) who filled out the WFC & FWC Scale at both time points (T1 and T2) were used to estimate the test-retest reliability.

## Measures

The Work-Family Conflict and Family-Work Conflict Scale (WFC & FWC Scale) (Netemeyer et al., 1996) consists of ten statements. Five statements describe conflicts in which the general demands of time devoted to and strain created by work interfere with family responsibilities (work-family conflict, WFC). Five statements describe conflicts in which family responsibilities interfere with work responsibilities (family work conflict, FWC) (Netemeyer et al., 1996). Participants were asked to indicate on a seven-point Likert scale whether the statements described their experiences (1 = *strongly disagree*, 7 = *strongly agree*). Higher mean scores indicated stronger WFC (Cronbach's  $\alpha = 0.88-0.89$ ) and FWC (Cronbach's  $\alpha = 0.83-0.89$ ) (Netemeyer et al., 1996).

Participants reported how much stress is caused (1 = *produces no stress*, 5 = *produces a great deal of stress*) by each of eleven described situations at work (Cavanaugh et al., 2000). Six items assessed challenge stressors (Cronbach's  $\alpha = 0.87$ ), i.e., demands at work that cause stress but are associated with potential gains for the individual (e.g., number of projects and assignments) (Cavanaugh et al., 2000). Five items assessed hindrance stressors (Cronbach's  $\alpha = 0.75$ ), i.e., demands at work that cause stress but are rarely associated with any potential gains for the individual (e.g., paper work/bureaucratic demands) (Cavanaugh et al., 2000). In the current study, challenge stressors have been assessed with an internal consistency of 0.88 and hindrance stressors with an internal consistency of 0.67.

Participants responded on a five-point Likert scale (1 = *totally disagree*, 5 = *totally agree*) to five questions asking to what extent they had career ambitions, i.e., would like to improve performance or pursue a higher position at work (Meeussen and Van Laar, 2018). Higher mean scores indicated strong career ambitions (Cronbach's  $\alpha = 0.88$ ) (Meeussen and Van Laar, 2018). In the current study, the Career Ambition Scale had an internal consistency of 0.85.

The four-item version of the Couples Satisfaction Index (CSI-4) (Funk and Rogge, 2007) was used to assess participants' relationship satisfaction. One item used a seven-point Likert scale (1 = *Not at all true*, 7 = *Perfect*), whereas the other items used a six-point Likert scale for responses (1 = *Extremely unhappy*, 6 = *Completely true*). Higher sum scores indicate stronger relationship satisfaction (Cronbach's  $\alpha = 0.94$ ) (Funk and Rogge, 2007). In the current study, the internal consistency of the CSI-4 was 0.93.

Additionally, sociodemographic information was assessed with self-constructed questions about participants' gender, age, sexual orientation, relationship status, highest level of education, nationality, position at the medical university, and number of children. For each question, participants could choose among several response options and give a free text response (Komlenac et al., 2021). For the analysis, all sociodemographic variables were dichotomized.

## Statistical Analysis

Descriptive statistics, correlations and the most widely used measure of reliability, namely Cronbach's alpha (Deng and Chan, 2017), were calculated with SPSS, version 26.0 (IBM Corp., Armonk, NY, United States). Because McDonald's  $\omega$  (McDonald, 1999) is recommended as an alternative indicator of reliability, especially when loadings of items vary (Cho and Kim, 2015; Deng and Chan, 2017), McDonald's  $\omega$  was calculated as a measure of scale reliability. Hierarchical McDonald's  $\omega$  ( $\omega_h$ ) was calculated as a measure of the general factor's reliability (Zinbarg et al., 2005; Zinbarg et al., 2006; Revelle and Zinbarg, 2009; Chung et al., 2016). McDonald's  $\omega$ s were calculated with R (R Core Team, 2021), version 4.1.1, using the *psych* package (Revelle, 2018) and the *MBESS* package (Dunn et al., 2014; Kelley, 2018). MPlus, Version 8 (Muthén and Muthén, 1998–2017) (Muthén & Muthén, Los Angeles, CA, United States), was used to calculate structural equation models. The fit indices of the structural equation models were calculated with the robust maximum likelihood (MLR) estimator (Kline, 2015; Muthén and Muthén, 1998–2017; Yuan and Bentler, 2000), because Kolmogorov-Smirnov tests revealed that variables were not normally distributed ( $Ds(274) > 0.07$ ,  $ps \leq 0.001$ ) (Field, 2009). The following cut-offs for the fit indices were used:  $p$ -value of scaled  $\chi^2 > 0.05$  (Weiber and Mühlhaus, 2014), root mean square error of approximation (RMSEA)  $\leq 0.08$  (Browne and Cudeck, 1993), standardized root mean square residual (SRMR)  $\leq 0.1$ , CFI  $\geq 0.90$ , and TLI  $\geq 0.90$  (Weiber and Mühlhaus, 2014).

Similar to the original and other validation studies of the WFC & FWC Scale (Netemeyer et al., 1996; Boyar et al., 2006), a one-factor model (all items loaded on a single factor) and a two-factor model (all items loaded on their respective factor) were calculated. Furthermore, in order to test whether items measure two constructs that can be assumed to represent one broad central construct (i.e., a "general factor") a bifactor model (all items loaded on a general factor in addition to their respective factor) was tested (Chen et al., 2012; Reise, 2012; Rodriguez et al., 2016). For the model with best fit indices, configural, metric, scalar, and residual invariance were tested (Kline, 2015). Models were compared with scale-adjusted chi-square difference tests (Satorra and Bentler, 2001; Satorra and Bentler, 2010)<sup>2</sup> and changes in CFI ( $\Delta$ CFI). Non-significant chi-square difference tests and  $\Delta$ CFI  $\leq 0.01$  indicate equivalent models (Cheung and Rensvold, 2002). Additionally, bias-corrected bootstrap confidence intervals (bootstrap sample  $n = 1,000$ ) were calculated in order to analyze similarities of factor loadings and intercepts between women's and men's structural equation models (Cheung and Lau, 2012).

## RESULTS

### Descriptive Statistics

On average, participants did *rather not* experience WFC and most reported not experiencing FWC (Table 1), whereby men

<sup>2</sup><http://www.thestatisticalmind.com/calculators/SBChiSquareDifferenceTest.htm> (accessed 2021-09-06)

**TABLE 1** | Means, standard deviations, Cronbach's alphas, and McDonald's omegas by gender.

Scales	T1		T2		<i>r</i> <sub>test-retest</sub>	α	ω (95%CI)	Men			Women			<i>t</i> (272)	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				<i>M</i>	<i>SD</i>	ω (95%CI)	<i>M</i>	<i>SD</i>	ω (95%CI)		
WFC	3.4	1.6	3.3	1.7	0.73**	0.92	0.92 (0.90–0.94)	4.1	1.4	0.91 (0.84–0.95)	3.2	1.6	0.92 (0.90–0.94)	4.0**	0.6
FWC	2.4	1.3	2.3	1.4	0.79**	0.90	0.90 (0.87–0.93)	2.8	1.4	0.93 (0.89–0.96)	2.2	1.3	0.89 (0.85–0.92)	3.1*	0.4
WC	2.9	1.3	2.8	1.4	0.81**	0.92	0.92 (0.90–0.93)	3.5	1.2	0.90 (0.83–0.94)	2.7	1.6	0.91 (0.89–0.93)	4.1**	0.6

Scale scores ranged from 1 to 7. Higher scores indicate greater work-family conflict or family work conflict. α, Cronbach's alpha coefficient; ω, McDonald's ω; WFC, work-family conflict; FWC, family work conflict; WC, work-related conflict; T1, Time point 1; T2, Time point 2; ω, McDonald's ω. \**p* < 0.01, \*\**p* < 0.001.

experienced both kinds of conflict more often than did women (Table 1). Cronbach's alphas and McDonald's ωs indicated satisfactory internal consistencies (Ponterotto and Ruckdeschel, 2007; Table 1). The hierarchical McDonald's ω (ω<sub>h</sub>; McDonald, 1999) was 0.70 in the total sample, 0.65 in the male subsample, and 0.67 in the female subsample. Test-retest reliabilities were above 0.70 (Table 1). In women, the WFC & FWC Scale correlated with experienced work stress, but were unrelated to career ambitions or relationship satisfaction (Table 2). In men, the WFC & FWC Scale also correlated with experienced work stress, but not with career ambitions. Larger FWC went along with reduced relationship satisfaction in men (Table 2).

### Confirmatory Factor Analysis

The one-factor model did not satisfactorily fit the data (Table 3). The two-factor model and the bifactor model had satisfactory fit indices, whereas the bifactor model proved to have better fit indices (Table 3). In the bifactor model, all items significantly loaded to their respective factor (λ = 0.43 –0.67) and to the general factor (λ = 0.50 –0.76).

### Invariance Analysis

The bifactor model had acceptable fit indices in women and men. Additionally, the model of configural invariance had acceptable fit indices (Table 4). Metric invariance was supported because the model for metric invariance did not have poorer fit indices than did the configural model. Additionally, factor loadings in women and men were similar (d<sub>ifference</sub> = –0.2 – 0.3; 95%[lower limit (LL): –1.4 – –0.3; upper limit (UL) 0.3 – 1.6]) (Cheung and Lau, 2012). However, scalar invariance was not supported because the scalar

model had significantly poorer fit indices than did the metric model (Table 4). This conclusion was supported by the finding that all intercepts (except one) were significantly different in women and men (d<sub>ifference</sub> = –0.3 – 1.1; 95%[–0.2 – 0.6; 0.8 – 1.5]) (Cheung and Lau, 2012). Because scalar invariance was therefore not assumed, residual invariance was not tested (Kline, 2015).

## DISCUSSION

The current study is the first to report psychometric properties of a German-language version of the widely used Work-Family Conflict and Family-Work Conflict Scale (WFC & FWC Scale) (Netemeyer et al., 1996). Similar to the original English version of the WFC & FWC Scale, the German-language version of the scale is characterized by having high internal consistencies in women and men (Netemeyer et al., 1996; Ponterotto and Ruckdeschel, 2007). An indicator of construct validity of the German-language version of the WFC & FWC Scale is the good fit of data to the two-factor model, comparable to results of validation studies of the original version of the WFC & FWC Scale (Netemeyer et al., 1996). Additionally, the current study revealed that the bifactor model has an even better fit to the data, indicating that the WFC & FWC Scale measures “distinct but related forms of interrole conflict” (Netemeyer et al., 1996). The good fit of the bifactor model and the relatively high hierarchical McDonald's ω also support the calculation of a general score for the WFC & FWC Scale (Reise et al., 2013). Such a general score proved to have satisfactory internal consistency and test-retest reliability (Table 1) (Ponterotto and Ruckdeschel, 2007).

However, arguments against using such a general score offer past research that found that WFC and FWC can have different consequences (Amstad et al., 2011). Whereas WFC was found to be mostly linked to negative work-related outcomes, FWC is more likely associated with family related outcomes (Amstad et al., 2011). In line with this past research, the current study showed FWC but not WFC to be linked to relationship satisfaction in men. With those findings and found associations between the WFC & FWC Scale and work stress, the current study shows evidence for convergent validity (Netemeyer et al., 1996). Divergent validity was supported by the fact that no links were found between the WFC & FWC Scale and career ambitions (Ellinas et al., 2018).

The current study's invariance analysis revealed, similar to a meta-analysis of the WFC & FWC Scale, that configural

**TABLE 2** | Convergent and discriminant validity.

Scales	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) WFC				0.59***	0.43***	–0.11	0.08
(2) FWC				0.40**	0.46***	–0.34*	0.15
(3) WC				0.57***	0.53***	–0.27*	0.13
(4) Stress hindrance	0.61***	0.42***	0.60***				
(5) Stress challenge	0.46***	0.34***	0.46***				
(6) CSI	0.09	–0.10	0.00				
(7) Career ambition	0.03	–0.03	0.00				

Below diagonal correlations in women (*n* = 211) are reported; above diagonal correlations in men (*n* = 63) are reported. WFC, work-family conflict; FWC, family work conflict; WC, work-related conflict; CSI, Couples Satisfaction Index. \**p* < 0.050, \*\**p* < 0.010, \*\*\**p* < 0.001.

**TABLE 3** | Goodness-of-fit indices of the models ( $N = 274$ ).

Model	Fit indices						
	S-B $\chi^2$	df	Scaling correction	CFI	TLI	RMSEA [90% CI]	SRMR
(1) 1-Factor	374.7**	35	1.558	0.712	0.630	0.188 [0.171–0.206]	0.128
(2) 2-Factor	101.7**	36	1.340	0.944	0.930	0.082 [0.063–0.101]	0.146
(3) Bifactor	47.6*	28	1.327	0.983	0.973	0.051 [0.024–0.075]	0.058
Model comparison	$\Delta\chi^2$	df	$p$	$\Delta$ CFI	Conclusion		
1 vs. 2	71.1	1	<0.001	0.232	Prefer 2		
2 vs. 3	52.8	8	<0.001	0.039	Prefer 3		

The conclusion in the model comparison section is based on a joint consideration of  $\Delta\chi^2$  and  $\Delta$ CFI. Chi-square statistics were estimated with the robust maximum likelihood (MLR) estimator and  $\Delta\chi^2$  was calculated with the scale-adjusted chi-square difference test (<http://www.thestatisticalmind.com/calculators/SBChiSquareDifferenceTest.htm>, accessed 2021-09-06). df, degrees of freedom; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; TLI, Tucker-Lewis Index; CFI, comparative fit index; CI, confidence interval. \* $p < 0.05$ , \*\* $p < 0.001$ .

**TABLE 4** | Measurement invariance across women and men.

Model	Fit indices						
	S-B $\chi^2$	df	Scaling correction	CFI	TLI	RMSEA [90% CI]	SRMR
Bifactor men	27.4	28	0.988	1.000	1.003	0.000 [0.00–0.095]	0.044
Bifactor women	49.6*	28	1.261	0.974	0.959	0.046 [0.061–0.088]	0.050
(1) Configural	79.7*	56	1.124	0.981	0.970	0.056 [0.023–0.082]	0.049
(2) Metric	86.7	73	1.190	0.989	0.986	0.037 [0.000–0.064]	0.067
(3) Scalar	115.2*	83	1.170	0.974	0.972	0.053 [0.026–0.075]	0.102
Model comparison	$\Delta\chi^2$	df	$P$	$\Delta$ CFI	Conclusion		
1 vs. 2	9.7	17	0.917	0.008	Equivalent		
2 vs. 3	30.9	10	<0.001	0.015	Not equivalent		

The conclusion in the model comparison section is based on a joint consideration of  $\Delta\chi^2$  and  $\Delta$ CFI. Chi-square statistics were estimated with the robust maximum likelihood (MLR) estimator and  $\Delta\chi^2$  was calculated with the scale-adjusted chi-square difference test (<http://www.thestatisticalmind.com/calculators/SBChiSquareDifferenceTest.htm>, accessed 2021-09-06). df, degrees of freedom; RMSEA, root mean square error of approximation; SRMR, standardized root mean square residual; TLI, Tucker-Lewis Index; CFI, comparative fit index; CI, confidence interval. \* $p < 0.05$ .

and metric invariance can be assumed across women and men (Shockley et al., 2017). Those results support the conclusion that the two scales making up the WFC & FWC Scale have similar meaning across women and men (Gregorich, 2006). However, current and previous findings (Shockley et al., 2017) suggest the presence of a differential additive (acquiescence) response style (Cheung and Rensvold, 2000; Kline, 2015). The finding that scalar invariance cannot be assumed and that all intercepts (except one) were significantly different in women and men (Cheung and Lau, 2012) indicates that one group is systematically giving higher or lower responses than the other group. Hence, a given score may not represent the same level of a latent factor in women and men. This affects the general score and the two scale scores (WFC and FWC). Thus, the currently found gender differences and past reported gender differences (e.g., Shockley et al., 2017) need to be interpreted with caution, because differences in mean scores may not represent differences on the (underlying) latent level of experienced WFC or FWC.

The current study is one of the first studies to report test-retest reliabilities of the WFC & FWC Scale (Min et al., 2021). Obtained estimates of the test-retest reliabilities for the German version of the WFC & FWC Scale can be interpreted as being acceptable (Nunnally and Bernstein, 1994; Vilagut, 2014) and exceed test-retest reliabilities reported for similar instruments assessing WFC

or FWC (Fisher et al., 2016; Min et al., 2021). Thus, the German-language version of the WFC & FWC Scale can be recommended for future longitudinal studies that analyze (systematic) changes in WFC or FWC over time (Borgmann et al., 2020).

The generalizability of the current study's findings is limited because the study's sample consisted of employees of only one organization. Even though lower bound sample size calculations and recommendations concerning minimum sample sizes justify the adequacy of the current study's sample size, this relatively small sample size further limits the generalizability of the findings (Barrett, 2007; Westland, 2010). Small sample sizes do not only reduce the trustworthiness of results, but also increase the risk for technical problems during the analysis (Kline, 2015). Thus, future studies of the WFC & FWC Scale should include larger samples. Future studies employing larger and more diverse samples are needed to test measurement invariance across groups additional to women and men, such as groups with different cultural background, age, sexual orientation, or gender identity (including participants who identify with a gender identity other than *cis*-women and *cis*-men). In the current study not only the WFC & FWC Scale was translated from English to German, but also the other instruments that were used to estimate convergent and divergent validity were available only in English and needed to be translated for the current study. Those

other instruments have not been subject to the same rigorous analyses of psychometric properties as was the WFC & FWC Scale. Thus, results concerning the convergent and divergent validity need to be interpreted with caution until future studies report on the psychometric properties of all the currently used instruments. Another limitation of the study is the absolutist approach, which is based on the assumption that a translated measure will be identical to the original version in all aspects (Swami and Barron, 2019). However, during translation more rigorous and more culture-specific adaptations might be needed for the translated version to be conceptually equivalent in a cultural context for which the original version was not developed (Swami and Barron, 2019). After the translation of items pre-testing in a relatively small sample can be conducted to assess with quantitative and qualitative methods the understanding, acceptability, and emotional impact of the culturally adapted items (Swami and Barron, 2019). Finally, as is the case with most questionnaire studies, the current study is based on self-reports and is therefore susceptible to socially desirable responding.

Despite its limitations the current study offers adequate evidence for validity and reliability of a German-language version of the Work–Family Conflict and Family–Work Conflict Scale for the assessment of two distinct but related constructs, namely WFC and FWC (Netemeyer et al., 1996). The found psychometric properties of the WFC & FWC Scale also support the calculation of a general score for future studies that do not need to distinguish between WFC and FWC. Mean differences between women and

men should be interpreted with caution because scalar invariance cannot be assumed. Future studies might improve the scale further by analyzing cultural appropriateness and developing items that are especially relevant and normative for specific target populations (Swami and Barron, 2019).

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

NK, LS, and MH designed the research. NK and MH collected the data. NK analyzed and interpreted the data and wrote the manuscript. TV critically commented the manuscript. All authors read and approved the final manuscript.

## REFERENCES

- Amstad, F. T., Meier, L. L., Fasel, U., Elfering, A., and Semmer, N. K. (2011). A meta-analysis of work–family conflict and various outcomes with a special emphasis on cross-domain versus matching-domain relations. *J. Occup. Health Psychol.* 16, 151–169. doi: 10.1037/a0022170
- Barrett, P. (2007). Structural equation modelling: adjudging model fit. *Personal. Individual Differ.* 42, 815–824. doi: 10.1016/j.paid.2006.09.018
- BGBI (2002). *Bundesgesetz über die Organisation der Universitäten und ihre Studien (Universitätsgesetz 2002 – UG) [Federal Act on the Organisation of Universities and their Studies]*, BGBI. I Nr. 120/2002. Available online at: [https://www.ris.bka.gv.at/Dokumente/BgblPdf/2002\\_120\\_1/2002\\_120\\_1.pdf](https://www.ris.bka.gv.at/Dokumente/BgblPdf/2002_120_1/2002_120_1.pdf)
- Borgmann, L.-S., Rattay, P., and Lampert, T. (2019). Health-related consequences of work–family conflict from a European perspective: results of a scoping review. *Front. Public Health* 7:189. doi: 10.3389/fpubh.2019.00189
- Borgmann, L.-S., Rattay, P., and Lampert, T. (2020). Longitudinal analysis of work-to-family conflict and self-reported general health among working parents in Germany. *Int. J. Environ. Res. Public Health* 17:3966. doi: 10.3390/ijerph17113966
- Boyar, S. L., Carson, C. M., Mosley, D. C., Maertz, C. P., and Pearson, A. W. (2006). Assessment of the validity of Netemeyer et al.'s (1996) WFC and FWC scales. *Int. J. Conflict Manag.* 17, 34–44. doi: 10.1108/10444060610734163
- Brislin, R. W. (1970). Back-translation for cross-cultural research. *J. Cross-Cultural Psychol.* 1, 185–216. doi: 10.1177/135910457000100301
- Browne, M. W., and Cudeck, R. (1993). “Alternative ways of assessing model fit,” in *Testing Structural Equation Models*, eds K. A. Bollen and J. S. Long (London: Sage Publications).
- Bundesgesetz über Krankenanstalten und Kuranstalten (KAKuG) (2016). *[Hospitals and Health Resorts Act]*, BGBI. I Nr. 3/2016. Available online at: <https://www.ris.bka.gv.at/eli/bgbl/I/2016/3>
- Carlson, D. S., Kacmar, K. M., and Williams, L. J. (2000). Construction and initial validation of a multidimensional measure of work–family conflict. *J. Vocational Behav.* 56, 249–276. doi: 10.1006/jvbe.1999.1713
- Cavanaugh, M. A., Boswell, W. R., Roehling, M. V., and Boudreau, J. W. (2000). An empirical examination of self-reported work stress among US managers. *J. Appl. Psychol.* 85, 65–74. doi: 10.1037/0021-9010.85.1.65
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., and Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: a comparison of the bifactor model to other approaches. *J. Pers.* 80, 219–251. doi: 10.1111/j.1467-6494.2011.00739.x
- Cheung, G. W., and Lau, R. S. (2012). A direct comparison approach for testing measurement invariance. *Organ. Res. Methods* 15, 167–198. doi: 10.1177/1094428111421987
- Cheung, G. W., and Rensvold, R. B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *J. Cross-Cultural Psychol.* 31, 187–212. doi: 10.1177/0022022100031002003
- Cheung, G. W., and Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Struct. Equat. Model. Multidisciplinary J.* 9, 233–255. doi: 10.1207/s15328007sem0902\_5
- Cho, E., and Kim, S. (2015). Cronbach's coefficient alpha: well known but poorly understood. *Organ. Res. Methods* 18, 207–230. doi: 10.1177/1094428114555994
- Chung, C. M., Liao, X., Song, H., and Lee, T. (2016). Bifactor approach to modeling multidimensionality of physical self-perception profile. *Measurement Phys. Educ. Exerc. Sci.* 20, 1–15. doi: 10.1080/1091367X.2015.1081594
- Deng, L., and Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educ. Psychol. Measurement* 77, 185–203. doi: 10.1177/0013164416658325
- Döring, N., and Bortz, J. (2016). “Psychologischer test,” in *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*, 5th Edn, eds N. Döring and J. Bortz (Berlin: Springer).
- Dunn, T. J., Baguley, T., and Brunsden, V. (2014). From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br. J. Psychol.* 105, 399–412. doi: 10.1111/bjop.12046
- Ellinas, E. H., Fouad, N., and Byars-Winston, A. (2018). Women and the decision to leave, linger, or lean in: predictors of intent to leave and aspirations to leadership

- and advancement in academic medicine. *J. Women's Health* 27, 324–332. doi: 10.1089/jwh.2017.6457
- Field, A. (2009). *Discovering Statistics Using SPSS*. London: Sage Publications Ltd.
- Fisher, G. G., Matthews, R. A., and Gibbons, A. M. (2016). Developing and investigating the use of single-item measures in organizational research. *J. Occup. Health Psychol.* 21, 3–23. doi: 10.1037/a0039139
- Funk, J. L., and Rogge, R. D. (2007). Testing the ruler with item response theory: increasing precision of measurement for relationship satisfaction with the couples satisfaction index. *J. Fam. Psychol.* 21, 572–583. doi: 10.1037/0893-3200.21.4.572
- Greenhaus, J. H., and Beutell, N. J. (1985). Sources of conflict between work and family roles. *Acad. Manag. Rev.* 10, 76–88. doi: 10.5465/amr.1985.4277352
- Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? testing measurement invariance using the confirmatory factor analysis framework. *Med. Care* 44, S78–S94. doi: 10.1097/01.mlr.0000245454.12228.8f
- Grzywacz, J. G., Frone, M. R., Brewer, C. S., and Kovner, C. T. (2006). Quantifying work–family conflict among registered nurses. *Res. Nurs. Health* 29, 414–426. doi: 10.1002/nur.20133
- Huang, J. L., Curran, P. G., Keeney, J., Poposki, E. M., and Deshon, R. P. (2012). Detecting and deterring insufficient effort responding to surveys. *J. Bus. Psychol.* 27, 99–114. doi: 10.1007/s10869-011-9231-8
- Kelley, K. (2018). *MBESS: The MBESS R Package R package Version 4.4.3*. Available online at: <https://CRAN.R-project.org/package=MBESS>.
- Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling*. London: Guilford Publications.
- Komlenac, N., Stockinger, L., and Hochleitner, M. (2021). *Family Supportive Supervisor Behaviors Moderate Associations Between Work Stress and Exhaustion: Testing the Job Demands – Resources Model in Academic Staff at an Austrian Medical University [Manuscript submitted for publication]*. Gender Medicine Unit, Medical University of Innsbruck.
- MacCallum, R. C., and Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annu. Rev. Psychol.* 51, 201–226. doi: 10.1146/annurev.psych.51.1.201
- McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Meeussen, L., and Van Laar, C. (2018). Feeling pressure to be a perfect mother relates to parental burnout and career ambitions. *Front. Psychol.* 9:2113. doi: 10.3389/fpsyg.2018.02113
- Min, H., Matthews, R. A., Wayne, J. H., Parsons, R. E., and Barnes-Farrell, J. (2021). Psychometric evaluation of work–family conflict measures using classic test and item response theories. *J. Bus. Psychol.* 36, 117–138. doi: 10.1007/s10869-019-09656-5
- Muthén, L. K., and Muthén, B. O. (1998–2017). *Mplus User's Guide*. Los Angeles, CA: Muthén & Muthén.
- Netemeyer, R. G., Boles, J. S., and McMurrin, R. (1996). Development and validation of work–family conflict and family–work conflict scales. *J. Appl. Psychol.* 81, 400–410. doi: 10.1037/0021-9010.81.4.400
- Nunnally, J. C., and Bernstein, I. H. (1994). *Psychometric Theory*. London: McGraw-Hill, Inc.
- Ponterotto, J. G., and Ruckdeschel, D. E. (2007). An overview of coefficient alpha and a reliability matrix for estimating adequacy of internal consistency coefficients with psychological research measures. *Percept. Mot. Skills* 105, 997–1014. doi: 10.2466/pms.105.3.997-1014
- R Core Team (2021). *R: A Language and Environment for Statistical Computing 4.0.4*. Vienna: R Core Team.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behav. Res.* 47, 667–696. doi: 10.1080/00273171.2012.715555
- Reise, S. P., Bonifay, W. E., and Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *J. Pers. Assess.* 95, 129–140. doi: 10.1080/00223891.2012.725437
- Revelle, W. (2018). *psych: Procedures for Psychological, Psychometric, and Personality Research R Package Version 1.8.4*. Available online at: <https://CRAN.R-project.org/package=psych>.
- Revelle, W., and Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: comments on Sijtsma. *Psychometrika* 74, 145–154. doi: 10.1007/s11336-008-9102-z
- Rodriguez, A., Reise, S. P., and Haviland, M. G. (2016). Evaluating bifactor models: calculating and interpreting statistical indices. *Psychol. Methods* 21, 137–150.
- Satorra, A., and Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66, 507–514. doi: 10.1007/bf02296192
- Satorra, A., and Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test Statistic. *Psychometrika* 75, 243–248. doi: 10.1007/s11336-009-9135-y
- Schnell, R., Bachteler, T., and Reiher, J. (2010). Improving the use of self-generated identification codes. *Eval. Rev.* 34, 391–418. doi: 10.1177/0193841x10387576
- Shah, R., and Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: looking back and forward. *J. Operat. Manag.* 24, 148–169. doi: 10.1016/j.jom.2005.05.001
- Shockley, K. M., Shen, W., DeNunzio, M. M., Arvan, M. L., and Knudsen, E. A. (2017). Disentangling the relationship between gender and work–family conflict: an integration of theoretical perspectives using meta-analytic methods. *J. Appl. Psychol.* 102, 1601–1635. doi: 10.1037/apl0000246
- Sirgy, M. J., and Lee, D.-J. (2018). Work-life balance: an integrative review. *Appl. Res. Qual. Life* 13, 229–254. doi: 10.1007/s11482-017-9509-8
- Swami, V., and Barron, D. (2019). Translation and validation of body image instruments: challenges, good practice guidelines, and reporting recommendations for test adaptation. *Body Image* 31, 204–220. doi: 10.1016/j.bodyim.2018.08.014
- Vilagut, G. (2014). “Test-retest reliability,” in *Encyclopedia of Quality of Life and Well-Being Research*, ed. A. C. Michalos (Dordrecht: Springer).
- Weiber, R., and Mühlhaus, D. (2014). *Strukturgleichungsmodellierung: Eine Anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. Berlin: Springer-Verlag.
- Westland, C. J. (2010). Lower bounds on sample size in structural equation modeling. *Electron. Commerce Res. Appl.* 9, 476–487. doi: 10.1016/j.elerap.2010.07.003
- Yuan, K.-H., and Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociol. Methodol.* 30, 165–200. doi: 10.1111/0081-1750.00078
- Zinbarg, R. E., Revelle, W., Yovel, I., and Li, W. (2005). Cronbach's  $\alpha$ , Revelle's  $\beta$ , and McDonald's  $\omega$  H: their relations with each other and two alternative conceptualizations of reliability. *Psychometrika* 70, 123–133. doi: 10.1007/s11336-003-0974-7
- Zinbarg, R. E., Yovel, I., Revelle, W., and McDonald, R. P. (2006). Estimating generalizability to a latent variable common to all of a scale's indicators: a comparison of estimators for  $\omega$ h. *Appl. Psychol. Measurement* 30, 121–144. doi: 10.1177/0146621605278814

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Komlenac, Stockinger, Vogler and Hochleitner. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.