

Determining the dimensionality and gender invariance of the MACE work-to-family enrichment scale using bifactor and approximate invariance tests



Authors:

Pieter Schaap¹
Eileen Koekemoer¹

Affiliations:

¹Department of Human Resource Management, Faculty of Economic and Management Sciences, University of Pretoria, Pretoria, South Africa

Corresponding author:

Pieter Schaap,
pieter.schaap@up.ac.za

Dates:

Received: 25 June 2020
Accepted: 10 Nov. 2020
Published: 28 Jan. 2021

How to cite this article:

Schaap, P., & Koekemoer, E. (2021). Determining the dimensionality and gender invariance of the MACE work-to-family enrichment scale using bifactor and approximate invariance tests. *SA Journal of Industrial Psychology/SA Tydskrif vir Bedryfsielkunde*, 47(0), a1821. <https://doi.org/10.4102/sajip.v47i0.1821>

Copyright:

© 2021. The Authors.
Licensee: AOSIS. This work is licensed under the Creative Commons Attribution License.

Read online:



Scan this QR code with your smart phone or mobile device to read online.

Orientation: Uncertainty about which measurement model of the MACE work-to-family enrichment scale (MACE-W2FE) is best supported by the data called for clarification.

Research purpose: The main aim of our study was to get clarity on the dimensionality of the MACE-W2FE. The secondary aim was to test for approximate invariance of the measure for gender groups.

Motivation for the study: Variations in the reported measurement models for the MACE-W2FE between studies are not conducive for theory development and called for clarification. Previous models reported were a multidimensional model and a second-order model. Approximate measurement invariance is a prerequisite for study differences between gender groups.

Research approach/design and method: We did seek to resolve the problem by using bifactor model analysis, factor strength indices and local indicator misspecification analyses using a sample of 786 South African employees. Invariance was tested using the alignment optimisation method.

Main findings: In this study, we solved a substantive research problem by determining that the data from the study best supported a single breadth factor or first-order factor model that was essentially unidimensional. The invariance tests across gender groups confirmed approximate configural, measurement and scalar invariances for the unidimensional model.

Practical/managerial implications: Researchers and practitioners may include the MACE-W2FE in studies as a single-aggregated score without negligible loss in measurement precision.

Contribution/value-add: The extended confirmatory factor analyses we conducted proved valuable in resolving the MACE-W2FE's dimensionality vacillations, thereby enhancing the validity of inferences made from scale scores.

Keywords: bifactor analysis; gender invariance; scale dimensionality; factor strength indices; local indicator misfit analysis; approximate invariance testing; MACE work-to-family enrichment scale; work-family enrichment; method factors.

Introduction

Over the past few decades, work–family research has been dominated by the conflict perspective (Greenhaus & Beutell, 1985) according to which the fulfilment of multiple work and family roles leads to experiences of conflict and stress and their concomitant detrimental effects (Eby, Casper, Lockwood, Bordeaux, & Brinley, 2005). The conflict perspective has also been the focus of most work–family studies conducted in Africa (Dubihlela & Dhurup, 2013; Koekemoer, Mostert, & Rothmann, 2010; Mostert, 2011; Opie & Henn, 2013). However, because of the growing attention given to positive psychology, international work–family researchers have come to realise that resources may be generated when multiple roles are occupied, resulting in positive outcomes for employees, organisations and families (Greenhaus & Powell, 2006; Voydanoff, 2002). As a result, international scholars, organisations and human resource practitioners increasingly focus on the positive aspects of the work–family interface. Nevertheless, the number of studies emphasising this positive interaction between work and family within the South African context is limited (De Klerk, Nel, Hill, & Koekemoer, 2013; Jaga, Bagraim, & Williams, 2013).

The most comprehensive framework for and most cited explanation of this positive linkage are provided by Greenhaus and Powell (2006) who define work–family enrichment (WFE) as the

'extent to which experiences in one role improve the quality of life (namely performance and affect) in the other role' (p. 73). The main premise of their theory is that the generation of resources is a crucial driver for the enrichment process and that resources can be transferred from one domain to another, resulting in increased performance and affect in the receiving role (Greenhaus & Powell, 2006).

Based on this well-known model of Greenhaus and Powell (2006), Carlson, Kacmar, Wayne and Grzywacz (2006) developed their work–family enrichment scale (WFES), which, although widely used, has been criticised for not reflecting all the facets of resources that the WFE model proposes and for containing double-barrelled items (i.e. conveying different elements instead of single ideas) (Carlson, Grzywacz, & Zivnuska, 2009).

To improve on this well-known international instrument, WFES, De Klerk et al. (2013) developed the MACE WFE instrument using a South African sample and obtained initial validation for their instrument (MACE is an acronym for the names of the authors). The MACE WFE instrument consists of two distinct bidirectional scales that can be used independently of each other, namely the MACE work-to-family enrichment scale (MACE-W2FE) and the MACE family-to-work enrichment scale (MACE-F2WE). The distinction made between the two bidirectional scales is consistent with international WFE literature (Carlson et al., 2006; Frone & Yardley, 1997). In this article, we focus on the more widely used MACE-W2FE.

De Klerk et al. (2013), based on their conceptualisation of the WFE construct as:

[T]he extent to which various resources from work and family roles have the capacity to encourage an individual and to provide positive experiences, and thereby enhance that individual's quality of life in the other role (i.e. performance and positive affect). (p. 4)

Included items in the MACE-W2FE that reflected four categories of resources gained, namely perspectives, affect, social capital and time management.

However, previous studies show the MACE-W2FE's four-dimensional model might not be sufficiently supported by the data as evident in the dimensionality variations of the MACE-W2FE reported across studies. De Klerk, Nel and Koekemoer (2015) and Van Zyl (2020) reported that data supported a correlated four-dimensional factor model, whereas in other studies (Koekemoer, Strasheim, & Cross, 2017; Marais, De Klerk, Nel, & De Beer, 2014), a correlated four-dimensional measurement model was reported, but a good fitting second-order (SO) factor model was used to alleviate multicollinearity in the exogenous part of a structural equation model (SEM). The SO factor models showed that a strong common factor underlies the MACE-W2FE, which can be consistent with an approximate unidimensional factor model with trivial group-specific

factors or a general factor underlying substantive group-specific factors. Koekemoer et al. (2017) and Marias et al. (2014) did not indicate clearly which of the former or latter assumptions applied to the SO factor model reported. They argued that in the presence of multicollinearity, a good-fitting SO factor model justified the use of a single-aggregated variable in the exogenous part of an SEM model when supported in theory. However, according to Chen, West and Sousa (2006), the use of SO factor model often goes unchallenged or is glossed over in SEM studies and not helpful in resolving the dimensionality question. Yet, in another study, the MACE-W2FE subscale scores formed the manifest indicators to a single latent factor that was incorporated in an SEM model with external variables (Koekemoer, Olckers, & Nel, 2020). Similar dimensionality vacillations were reported for Carlson et al.'s (2006) WFES, indicating the possible existence of a common problem (Jiang & Men, 2017; Rastogi, Karatepe, & Mehmetoglu, 2018; Russo, Buonocore, Carmeli, & Guo, 2018; Siu et al., 2015; Timms et al., 2015).

Pertaining to the above-mentioned dimensionality issues, Garrido, González, Seva and Piera (2019) warn about treating substantively multidimensional scores as unidimensional (e.g. a single latent factor). Such factor scores are expected to lead to biased item parameter estimates and loss of information where they cannot be univocally interpreted. On the contrary, when factor scores can be univocally interpreted, treating the items as substantively multidimensional leads to factors of little theoretical interest and unclear interpretations. In conclusion, there is a clear need for clarity over the MACE-W2FE's dimensionality for it could impact negatively on the validity of score inferences and WFE theory development in general.

A bifactor model analysis can effectively resolve model dimensionality uncertainties because a bifactor model is theoretically consistent with a correlated first-order multidimensional factor model and a SO factor model (Rodríguez, Reise, & Haviland, 2016a). Where a strictly unidimensional model is rejected by model fit indices, bifactor analysis is useful in determining the strength of the general factor that underlies a multicomponent measure and the strength of each component after controlling for the common factor. A multidimensional measure may be assumed where one or more components show sufficient strength in terms of reliable variance. An approximate or essentially unidimensional (i.e. a single breadth factor) measure may be assumed to the extent that the factor score is univocal with ignorable biasing effects of the multidimensional components (Rodríguez et al., 2016a). Local item misspecification analysis allows for the evaluation of the extent to which misspecifications show ignorable biasing effects on the factor score of an assumed essentially unidimensional model.

Furthermore, findings about differences in gender group experiences of WFE have been contradictory, and therefore

more research on the topic is much needed (Rothbard, 2001; Van Steenberghe, Ellemers, & Mooijaart, 2007). Moreover, gender studies require the MACE-W2FE to show at least approximate measurement invariance for gender groups.

We argue that the strong emphasis on 'golden rules' for goodness-of-fit proposed by Hu and Bentler (1999) in deciding model fit, in the absence of an in-depth analysis of the measurement model, is the likely reason for the different measurement models used in the WFES and MACE-W2FE studies (Greiff & Heene, 2017; McNeish, An, & Hancock, 2018; Ropovik, 2015). Solely relying on confirmatory factor analysis (CFA), goodness-of-fit indices without additional analyses has proved to be ineffective in determining the dimensionality of a measure (Rodriguez et al., 2016a). In order to contribute to existing WFE literature and resolve the MACE-W2FE dimensionality and invariance issues, we followed a:

[S]ubstantive-methodology synergy approach where methodological advances [i.e. in-depth analyses techniques and alignment optimisation] are applied to substantive areas of research in order to obtain more precise answers to complex questions. (Marsh & Hau, 2007, p. 152).

For our study, we formulated the following research questions: (1) Are the MACE-W2FE's subdimensions substantively unique constructs? or (2) Is the MACE-W2FE an essentially unidimensional construct? (3) Is the MACE-W2FE second-order model theoretically plausible and clearly interpretable? (4) Can the MACE-W2FE be considered an approximate invariant measure to use across gender groups?

In seeking answers to the research questions, we demonstrated the usefulness of what we called 'extended CFA analyses' which included the following: bifactor testing, local indicator misspecification analysis, and approximate measurement invariance testing.

Our study aimed to contribute to work–family literature by providing rigorous evidence relating to the dimensionality and scale invariance of the MACE-W2FE within a South African sample.

Firstly, we discuss the substantive issues regarding the WFE theoretical framework, the development of the MACE instrument and related validity evidence. Thereafter, we discuss the methodological issues of CFA in testing model dimensionality, the use of extended analysis to resolve the MACE-W2FE's dimensionality vacillations and approximate invariance testing.

Substantive issues: Theoretical background and the development of the MACE instrument

Theoretical background

In recent years, numerous researchers have shown interest in the measurement of WFE because of the realisation that

organisations stand to benefit from recognising and accommodating employees' work–life needs (Shockley & Singla, 2011). Various models or frameworks to explain WFE have been put forward and the most prominent theories on which they are based are the theory of role accumulation (Sieber, 1974), the resource–gain–development perspective (Wayne, Grzywacz, Carlson, & Kacmar, 2007) and the work–home resources model (Ten Brummelhuis & Bakker, 2012).

When considering the *theory of role accumulation*, the literature attempts to explain how participation in multiple roles can produce positive outcomes for individuals, by putting forth three notions. The first notion being that work experiences and family experiences can have additive effects on well-being. In this sense, the argument is being made that individuals who participate in – and are satisfied with – work and family roles experience greater well-being than those who are dissatisfied with one or more of their roles. The second view researchers use to describe role accumulation is the idea that participation in both work and family roles can buffer individuals from distress in one of the roles. This notion dates back to the work of Sieber (1974) which stated that individuals who accumulate roles may compensate for failure in one role by falling back on gratification in another role. The third explanation put forward for role accumulation is that the experiences in one role can produce positive experiences and outcomes in the other role. It is also this specific explanation which Greenhaus and Powell (2006) utilised when developing their well-cited model of WFE. According to these authors this third mechanism best captures the concept of WFE as 'the extent to which experience in one role improve the quality of life in the other role' (p. 73).

When considering the *resource–gain–development perspective* (Wayne et al., 2007), the basic premise is that individuals have natural tendencies to grow and develop. When individuals engage in a role, they obtain resources so that they can experience positive gains. When gains from one domain are applied, sustained and reinforced in another, the end results are improved system functioning or facilitation.

It is against this backdrop that Greenhaus and Powell (2006) developed their WFE model. Work–family researchers agree that this model is one of the most comprehensive and systematic models of all that explains within-domain and cross-domain effects (Zhang, Xu, Jin, & Ford, 2018). As mentioned earlier, the generation of resources is crucial in the enrichment process. The main premise of the WFE model of Greenhaus and Powell (2006) is that the resources acquired in one role can enrich the other role through instrumental and/or affective paths. According to Greenhaus and Powell, a resource is an asset that may be drawn on when needed to solve a problem or cope with a challenging situation. Their WFE model identifies five types of resources that can be generated in a role:

- *Skills and perspectives*: Skills refer to a broad set of task-related cognitive and interpersonal skills, coping skills, multi-tasking skills, and knowledge and wisdom derived from role experiences. Perspectives involve ways of perceiving or handling situations which, in short, allow one to expand one's 'world view'.
- *Psychological and physical resources*: These include positive self-evaluations such as self-efficacy, self-esteem, personal hardiness, positive emotions about the future (e.g. optimism and hope) and physical health.
- *Social-capital resources*: There are two social-capital resources – influence and information – and they are derived from interpersonal relationships in work and family roles that may assist individuals in achieving their goals.
- *Flexibility*: This refers to the discretion to determine the timing, pace and location of meeting role requirements.
- *Material resources*: These include money and gifts obtained whilst fulfilling work and family roles.

Various instruments were developed based on the Greenhaus and Powell's (2006) model, of which the bidirectional WFES (work-to-family direction and family-to-work direction) of Carlson et al. (2006) is the most widely used internationally, but it is criticised for encompassing only three of the resources of the original model, namely development, affect and capital. Against this backdrop, the MACE WFE instrument consisting of the MACE-W2FE and MACE-F2WE was developed utilising a South African sample.

The MACE work-to-family enrichment scale

De Klerk et al. (2013) followed a comprehensive and thorough process of developing the items for the MACE-W2FE's four dimensions: they generated, modified and evaluated the items as per DeVellis (2003) guidelines. The final version of the MACE-W2FE used in this study consisted of 18 items and four dimensions.

The dimensions included in the MACE-W2FE were defined as follows:

- *Perspectives (P)*: individuals' participation in the work role that leads to the acquisition or refinement of skills, perspectives and values that improve the individuals' quality of life in the family role.
- *Affect (A)*: individuals' participation in their work role that leads to the acquisition or refinement of self-concept, positive affect and increased energy levels, and mental sharpness that improves the individuals' quality of life within the family role.
- *Time management (TM)*: individuals' participation in the work role that provides the ability to determine the timing and pace at which role requirements are met that improves the individual's quality of life within the family role.
- *Social capital (SC)*: individuals' participation in the work role that leads to the acquisition or refinement of the

maintenance of relationships and support that improves the individual's quality of life within the family role.

De Klerk et al. (2013) found no differential item functioning between gender groups across the full set of items in the MACE-W2FE using Rasch modelling techniques. They also found that gender groups' mean score for all items on the MACE-W2FE did not differ.

Using CFA, De Klerk et al. (2015) evaluated the structural validity of a shortened 18-item version of the MACE-W2FE. This four-dimensional model showed a good model fit according to generally accepted conventional criteria ($\chi^2 = 364.31$, $p < 0.01$; comparative fit index [CFI] = 0.97; Tucker-Lewis index [TLI] = 0.96 and root mean square error of approximation [RMSEA] = 0.05) (Hu & Bentler, 1999). It also showed high correlations between dimensions ranging from 0.54 to 0.63, suggesting shared variance ascribed to a common factor. Koekemoer et al. (2017) and Marais et al. (2014) used a well-fitting SO factor model of the MACE-W2FE to avoid multicollinearity in their respective studies on the antecedents and outcomes of WFE amongst female and married workers. Studies by De Klerk et al. (2015) and Koekemoer et al. (2017) confirmed that outcomes of WFE, such as job satisfaction, commitment, subjective career success and work engagement, strongly related to the MACE-W2FE's SO factor model with correlations varying between 0.50 and 0.66. More modest correlations (0.26–0.43) with the MACE-W2FE's subscales and relevant constructs, such as job satisfaction, work vigour, work dedication and career satisfaction, were obtained (De Klerk et al., 2015). The reported evidence suggests that the MACE-W2FE may be a valid measure of the construct WFE for the South African samples that were used. However, the variances in the MACE's measurement model used in the respective studies are not conducive for WFE theory development and need to be resolved.

Greenhaus and Powell's (2006) definition of WFE suggests that WFE can be viewed as an essentially unidimensional or broad construct that is informed by events and outcomes across the full spectrum of WFE resources. Support for the assertion can be found in Kacmar, Crawford, Carlson, Ferguson and Whitten's (2014, p. 45) study, where both the original nine-item version and the shortened three-item version of Carlson et al.'s (2006) WFE (direction work-to-family enrichment) is depicted as a single latent variable consisting of items that represent a spectrum of WFE resources (i.e. development, affect and capital resources). However, it is unclear whether an essentially unidimensional model or a multidimensional model of the MACE-W2FE is best supported by data. We argue that by using extended CFA analyses we can provide rigorous evidence relating to the dimensionality and scale invariance of the shortened version (consisting of 18 items) of the MACE-W2FE when applied to a large South African sample.

Methodological issues: Extending confirmatory factor analysis for evaluating scale dimensionality

Dimensionality issues

Theory testing in the social sciences is commonly associated with testing competing CFA measurement models (Marsh & Hau, 2007; Strauss & Smith, 2009). Routinely, a unidimensional measure model is tested, followed by the testing of multidimensional models, as dictated by plausible theoretical conceptualisations of the construct of interest. However, accepting the results of CFA analyses at face value is potentially dangerous. For example, a one-factor model (see Figure 1a) containing numerous items and allowing large degrees of freedom hardly ever describes real data and is routinely rejected based on the results of statistical model fit indices (Bentler, 2009). When applying theory, the prospect of finding a perfectly unidimensional model in assessment data is nil (Reise, Moore, & Haviland, 2010). In contrast, when the same data are subjected to correlated first-order CFA models (see Figure 1b), the multidimensional model will almost always be supported (Reise et al., 2010). Correlated first-order factor models often deceptively show good model fit and salient group-specific factors. The deception of a good-fitting correlated first-order factor model is created by a substantive general factor running amongst all the items and by the differentiating effect of parallel item wording or method artefacts (Reise et al., 2010). Highly correlated group-specific factors in first-order factor models rarely reflect unique and substantive factor variance after partialling out the common variance from a substantive general factor (Rodriguez et al., 2016a). Cattell and Tsujioka (1964) argue that without skilful factor analysis, detecting pseudo-specific group factors consisting of narrow bloated specifics or systematic biases is hard. Bloated specifics with little substance are common occurrences in published scales and are difficult to detect (Rodriguez et al., 2016a).

Where the factors in multidimensional models correlate strongly, researchers often adopt a SO factor model (see Figure 1c) (Chen et al., 2006), especially when multicollinearity may be a concern, as was the case with the MACE studies in Koekemoer et al. (2017) and Marais et al. (2014). Gignac (2016) acknowledges that the SO (i.e. higher order) model is the only model where hypothesis with respect to the association between group-level factors and the general factor can be tested. However, a SO factor model constrains the first-order item loadings to be equal within each factor and is known as the proportionality constraint. Imposed proportionality constraints on items in a SO factor model may represent an unnatural and difficult-to-interpret model solution despite obtaining good model fit according to conventional standards (Gignac, 2016). Gignac (2016) argued that:

[E]mpirically and theoretically, researchers may find it difficult to explain why the nature of the general factor in a second-order

model is such that each and every item within a specific factor can contribute variance to the general factor and the specific factors' residual in a perfectly equal proportional manner. (p. 65)

Moreover, SO factor models do not give clear answers on the extent to which a measure is unidimensional versus multidimensional (Rodriguez et al., 2016a).

Unless researchers are mindful of the dimensionality issues that we have pointed out and the limitations of global fit indices, it may result in defective measurement models being accepted as close-to-fitting. Hayduk (2014) urges researchers to do a diagnostic assessment of models before accepting a model as sufficiently supported by the data. We now turn our attention to bifactor and local indicator misspecification analyses techniques that can be applied for diagnosing and resolving of dimensionality issues.

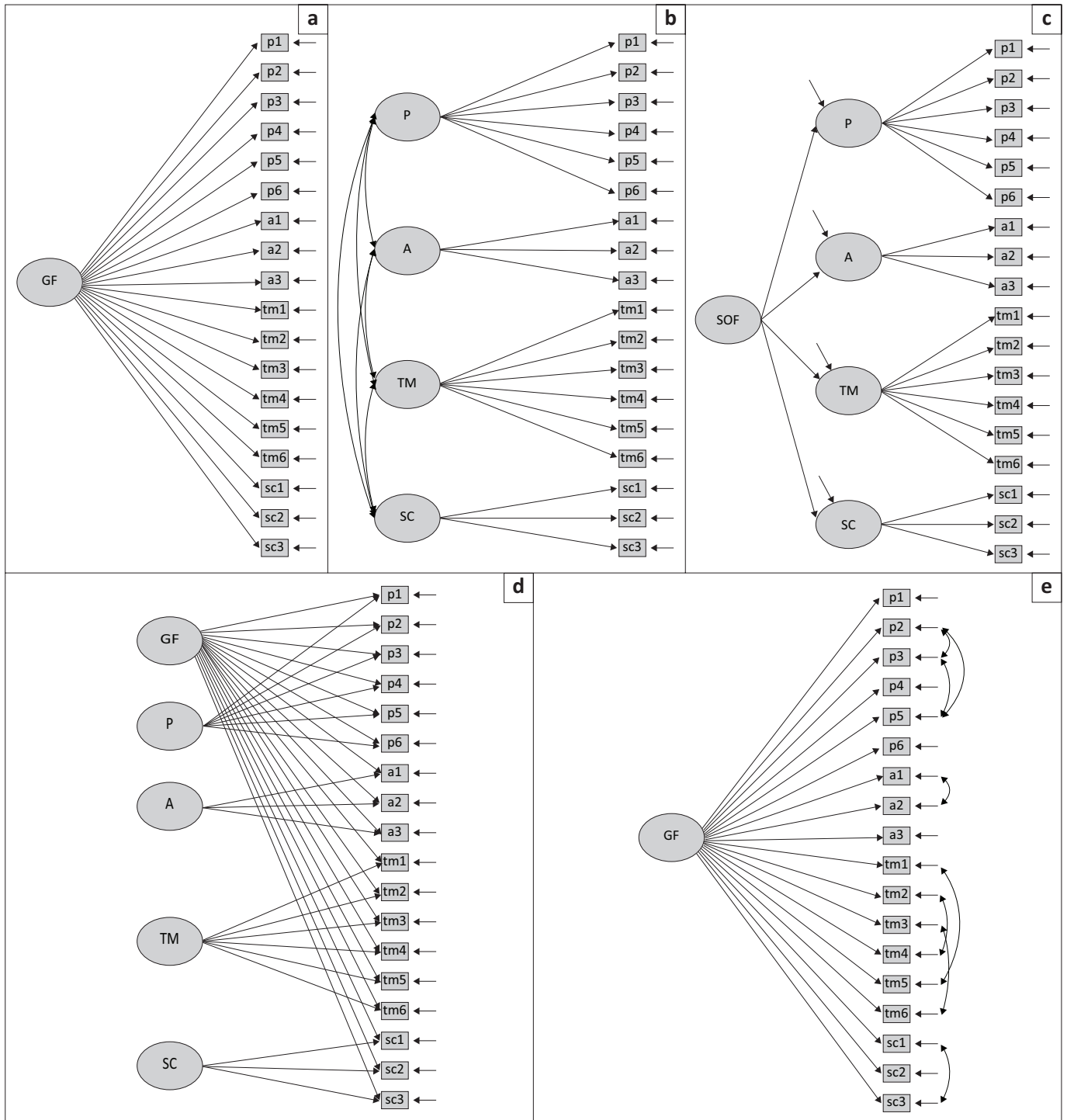
Analyses to resolve dimensionality issues

Bifactor analysis

Bifactor modelling (see Figure 1d) allows researchers to simultaneously investigate unidimensionality and multidimensionality by placing the common factor and group factors on an equal conceptual footing to compete for item variance (Reise et al., 2010). The bifactor model specifies that each item simultaneously explains a portion of a common factor and a portion of a single group factor (Reise et al., 2010). Bifactor modelling is an effective technique for resolving if a measure is essentially unidimensional or distinctly multidimensional. Obtaining clarity about the dimensionality of a measurement model can assist in avoiding multicollinearity problems when SEM or another form of multiple regression analysis with external variables is used (Rodriguez et al., 2016a).

Unlike SO factor models, the proportionality constraint does not apply in bifactor models and the item parameter estimates are freely estimated for both the general factor and the specific group factor. Where the data violate the proportionality constraint in the SO factor model, the bifactor model will always show a better fit that corresponds to the degree of violation (Gignac, 2016). Whereas the items in a SO factor model only indirectly affect the general factor via the specific group factor, the items in a bifactor model have a direct effect on both the general factor and the group factor.

Supporting bifactor strength indices (see the 'Methodology' section for the details on strength indices) can be applied to evaluate the extent to which the model supports essential unidimensionality and the plausibility of unique multidimensional factors after partialling out the common factors' variance (Reise et al., 2010; Rodriguez et al., 2016a; Rodriguez, Reise, & Haviland, 2016b). Unique multidimensional factors in bifactor models are also known as residualised factors (i.e. factors that show common variance after removal of the general factor's variance) (Rodriguez et al., 2016b).



GF, general factor; P, work–family perspectives; A, work–family affect; TM, work–family time management; SC, work–family social capital; SOF, second-order factor.

Note that the diagrams are only intended to be illustrative of the different model structures and providing larger labels would make models too large to present.

FIGURE 1: Models tested: (a) one-factor model, (b) four-factor model, (c) second-order factor model, (d) bifactor model and (e) one-factor model with method artefacts.

Local item misspecification analyses

Researchers warn against overreliance on simplistic global model fit indices when determining the dimensionality of measures in the social sciences (Greiff & Heene, 2017) because models in this field are always simplifications of reality because of the imperfect nature of data; consequently, these models are always misspecified to some extent (Saris, Satorra, & Van Der Veld, 2009). Therefore, it is important to

supplement global fit indices with local item misspecification analyses to avoid substantively irrelevant misspecifications leading to rejecting a model or overlooking substantively relevant misspecifications in model acceptance (Saris et al., 2009). According to Sellbom and Tellegen (2019), correlated residuals are very important sources of misspecification in CFA models and should be examined to avoid biased results when evaluating global model (mis)fit.

Approximate invariance testing

Proven measurement invariance and scalar invariance are prerequisites for making valid statistical conclusions about scale mean differences of groups under varied conditions (Sass, 2011). The alignment method for multiple-group CFA can be used to compare factor means and variance of groups without requiring exact measurement invariance (Asparouhov & Muthén, 2014). The conventional multiple-group CFA without the alignment method is inclined to be too strict in the identification of non-invariant parameters, leading to a series of model adaptations that may be data-specific or misspecified (De Bondt & Van Petegem, 2015). In the alignment method, measurement invariance is estimated without the need to constrain factor loadings and intercepts to being equal, for the optimal measurement invariant pattern is effectively discovered through alignment optimisation. The alignment optimisation procedure applies a simplicity function that works like the rotation criteria in exploratory factor analysis and retains the unrestricted configural model (model zero) but minimises non-invariance without compromising model fit. According to Asparouhov and Muthén (2014), up to 25% parameters may be non-invariant without adversely impacting on the reliable comparison of the factor means of groups. In other words, the alignment method does not require all differences in factor loadings (measurement invariance) and intercepts (scalar invariance) to be strictly zero before valid factor mean comparisons for groups can be made. The imperfect nature of item responses is a reality in the social sciences and this imperfection affects invariance and theory testing in SEM, but it can be accommodated through innovations such as the alignment optimisation method that allows for approximate measurement invariance (Asparouhov & Muthén, 2014).

Current study

In an attempt to stimulate future work–family studies from Africa, we investigated the dimensionality of the MACE W2FE instrument and gender invariance using extended CFA analysis techniques. Based on our literature discussion, we present the following hypothesis with respect to the South African sample surveyed:

H1a: We hypothesised an essentially unidimensional measurement model for the MACE-W2FE (see Figure 1a and 1e).

H1b: We further hypothesised that the multidimensional elements of the MACE-W2FE are not distinct and substantive constructs (see Figure 1b and 1d).

H2: We also hypothesised that the proportionality constraints for the SO factor model for the MACE-W2FE have been violated (see Figure 1c).

H3: Finally, we hypothesised that the MACE-W2FE will show approximate configural, measurement and scalar invariances for gender groups.

Furthermore, we demonstrated the use and value of bifactor (see Figure 1d) and local indicator misspecification analyses in resolving the MACE-W2FE's dimensionality vacillations and we tested approximate gender invariance at different levels of measurement using the alignment optimisation technique.

Methodology

Research design

Using a quantitative cross-sectional research design, we collected survey data to investigate the dimensionality, invariance and model specifications of the MACE-W2FE.

Research sample

Cross-sectional survey data were obtained from a convenience study sample ($N = 786$) of South African employees from industry sector such as mining, engineering, IT, manufacturing, finance and education. The majority of the sample consisted of Caucasian (86%) female employees (70%), of whom 67% was married, 85% had children and 14% was single. Of the sample, 50% possessed a degree or a postgraduate degree.

We used an anonymous web-based survey to obtain respondents' biographical information and to administer the MACE work–family instrument. We informed the participants that their participation was voluntary, and we obtained their informed consent. Ethical approval was obtained from the Research Ethics Committee of the relevant higher education institution.

The MACE-work-to-family enrichment scale

Our investigation concerned the 18-item MACE-W2FE developed by De Klerk et al. (2013). It includes the following four dimensions or subscales: *work–family perspectives* (six items relating to skills gained, e.g. 'My family life is improved by my work showing me different viewpoints'); *work–family affect* (three items relating to feelings gained, e.g. 'My family life is improved by work that makes me feel happy'); *work–family time management* (six items, e.g. 'My family life is improved by managing my pace at work') and *work–family social capital* (three items relating to the support participants receive from colleagues, e.g. 'My family life is improved by the support I receive from my colleagues'). A Likert-type rating scale ranging from 1 (strongly disagree) to 5 (strongly agree) was used. Previous studies (De Klerk et al., 2015; Marais et al., 2014; Van Zyl, 2020) showed acceptable Cronbach's alpha coefficients: work–family perspectives (ranging between 0.91 and 0.96), work–family affect (ranging between 0.84 and 0.95), work–family time management (ranging between 0.90 and 0.92) and work–family social capital (ranging between 0.80 and 0.87).

Analyses

We used the Mplus Statistical Software Version 8.3 and the maximum likelihood estimation method with robust standard errors (MLR) to test the measurement models included in this study. The MLR compensates for deviations from the multivariate normality assumption associated with Likert-type scales (Muthén & Muthén, 2017; Schmitt, 2011). To achieve the purpose of the study, we tested all the CFA models depicted within Figure 1, namely an essentially

unidimensional model (Figure 1a), a four-factor model (Figure 1b), a SO factor model (Figure 1c), a bifactor model (Figure 1d) and an essentially unidimensional model with method artefacts (Figure 1e). The model depicted in Figure 1a was used to test the gender invariance of the MACE-W2FE. We assessed the sample size adequacy for the purposes of the analyses using Kaiser–Meyer–Olkin measure of sampling adequacy (KMO) and Bartlett’s test of sphericity ($KMO > 0.70, p < 0.01$) (Cerny & Kaiser, 1977).

To evaluate the plausibility of the CFA models, we used the chi-square goodness-of-fit test ($\chi^2, p < 0.05$), the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA) and standardised root mean square residual (SRMR) as global indices of model fit. Model fit, according to CFI and TLI indices, is considered acceptable and good when exceeding 0.90 and 0.95, respectively. The RMSEA and SRMR values of less than 0.05 and 0.08, respectively, reflect a close fit and a reasonable fit to the data (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004). In addition, we used the Akaike information criterion (AIC) to compare alternative models, whereby the model with the lowest AIC value is the better model. As indicators of a significant difference in model fit where nested models are compared (Chen, 2007), we relied on changes greater than 0.01 on CFI, TLI and RMSEA, and a statistically significant ($p < 0.01$) adjusted χ^2 with the Satorra–Bentler scaling correction formula (Chen, 2007).

We adopted the notion that the results of global model (mis) fit indices are preliminary and require an evaluation of local parameter misspecifications (which are a source of model misfit) before final conclusions on model fit can be made (Marsh et al., 2004). We used Jrule software for Mplus to evaluate the local parameter misspecifications on the correlated residuals (Oberski, 2009), being the most important source of misspecification in measurement models. According to the Saris–Satorra–Van der Veld approach (Saris et al., 2009), the statistically overly sensitive modification indices (MI) should be considered alongside Cohen’s (1992) criterion for sufficient statistical power ($1 - \beta > 0.80$). Substantive local misspecification is evident in the presence of a statistically significant ($p < 0.05$) modification index and low statistical power ($1 - \beta < 0.80$). However, when the modification index is statistically significant ($p < 0.05$) and statistical power is high ($1 - \beta > 0.80$), the expected parameter changes (EPCs) for that indicator need to be outside the range of -0.10 to 0.10 , to be considered substantively relevant. In the latter case, where EPC is small (e.g. within the range of -0.10 to 0.10), it can be concluded that no relevant misspecification is prevalent that deviates substantively from zero.

We used the bifactor model to evaluate the distinctiveness of the specific or group factors and the plausibility of an essential general factor for the MACE-W2FE. The bifactor analysis of measures is a good choice where both correlated factors and SO CFA models show a good fit

(Reise, 2012). To determine whether the data sufficiently supported a distinct first-order group-factor model or whether a unidimensional model could be assumed, we used a variety of factor strength indices applicable for evaluating bifactor models (Reise et al., 2010; Reise, Bonifay, & Haviland 2013; Reise, Scheines, Widaman, & Haviland, 2013; Rodriguez et al., 2016a, 2016b). Detailed definitions, formulas and discussions of the factor strength indices are beyond the scope of this article and are available in Reise et al. (2010, 2013) and Rodriguez et al. (2016a, 2016b). These indicators were the following: explained common variance (ECV); McDonald’s (1999) omega reliabilities; omega (ω), omega hierarchical (ω_H/ω_{HS}), construct replicability (H), factor determinacy (FD) and percentage of uncontaminated correlations (PUCs). We used the absolute average relative parameter bias (ARPB) index at factor level and ARPB-I at item level to evaluate bias on factor loadings attributed to factor misspecifications.

An ARPB of below 10%–15% between the factor loadings of the common factor of a bifactor model and a unidimensional model can be considered non-substantive suggesting an essentially unidimensional model where ECV, PUC and ω_H values are 0.70 or higher. Percentage of uncontaminated correlations moderates ECV when considered concurrently. When PUC is high (> 0.80) and ECV is as low as 0.50, essential unidimensionality may still apply. However, when PUC is lower (< 0.80), ECV should be greater than 0.60 (e.g. PUC = 0.70 and ECV = 0.70) and ω_H should be greater than 0.70 to assume essential unidimensionality. Where H and FD^2 are equivalent and exceed 0.80, its essential unidimensionality is supported. Factor determinacy should exceed 0.90 before the use of factor scores instead of latent variables in an SEM model is justified and H exceeding 0.80 suggests good factor replicability. However, H and FD can be bloated by very narrow factors or bloated specifics and should be interpreted with caution. H and FD^2 values exceeding 0.70 could signify plausible group factors or subscales. A minimum value of 0.50 and preferably closer to 0.75 for ω_{HS} suggest a substantive group factor and multidimensionality.

By comparing the nested bifactor model to the SO factor model, we determined if proportionality constraints had been violated in the SO factor model, and we relied on model change statistics to confirm significant violations (Yung, Thissen, & McLeod, 1999).

Finally, we estimated the approximate invariance of the MACE-W2FE for gender groups using MLR estimation and the alignment optimisation method (Asparouhov & Muthén, 2014). In alignment optimisation, the configural invariance model is used as the baseline model. Next, we conducted the factor loading and intercept invariance tests where the total amount of non-invariance is minimised using a simplicity function for every pair of groups and for every intercept and loading using a component loss function from EFA rotations (Muthén & Asparouhov, 2018).

Ethical consideration

The approval is subject to the researcher abiding by the principles and parameters set out in the application and research proposal in the actual execution of the research. The approval does not imply that the researcher is relieved of any accountability in terms of the Codes of Research Ethics of the University of Pretoria if action is taken beyond the approved proposal. If during the course of the research it becomes apparent that the nature and/or extent of the research deviates significantly from the original proposal, a new application for ethics clearance must be submitted for review.

Results

In this section, we present a summary of the descriptive statistics and the results of estimating the CFA models (one-factor, four-factor, second-order factor and bifactor).

Descriptive statistics

The descriptive statistics showed item scores that varied between 3.5 and 3.9, the average score being 3.70. The standard deviations varied between 0.73 and 1.08; the mean deviation being 0.83. The item skewness varied between -1.01 and -0.50; the mean skewness being -0.72. The item kurtosis varied between -0.40 and 1.3; the mean being 0.35. The data signified a good approximation of the normal distribution (skewness and kurtosis between -1 and +1). The KMO and Bartlett's test of sphericity for sample size adequacy were, respectively, 0.94 and $p < 0.000$. Therefore, the sample size was considered adequate (KMO > 0.70 , $p < 0.01$) to continue with the CFA analyses.

Estimated confirmatory factor analysis models

As shown in Table 1, all the global fit indices (i.e. CFI, TLI, RMSEA and SRMR) did not support a unidimensional (one-factor) model when using the golden rules for model fit (Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004). The one-factor structure given in Table 2 can be considered well defined ($\lambda = 0.63$ – 0.81 ; mean (M) = 0.70). Cronbach's alpha reliability for the one-factor model was 0.94. However, the correlated four-factor model showed a good fit on all the indices, but the sample size sensitive χ^2 was significant ($p < 0.01$). As indicated in Table 2, we obtained a well-defined factor structure for the four-factor model with overall high loadings ($\lambda = 0.67$ – 0.89 ; $M = 0.80$). However, the factor correlation matrix showed high correlations between factors ($r = 0.66$ – 0.75 ; $M = 0.70$), suggesting that a common factor underlies the model. The Cronbach's alpha reliabilities for the four-factor model were all high ($P = 0.92$, $A = 0.80$, $TM = 0.90$ and $SC = 0.83$) suggesting high item homogeneity and narrow scales.

The SO factor model showed a good fit for all the indices, but the χ^2 was significant ($p < 0.01$). The SO factor loadings were high (SO: $\lambda = 0.81$ – 0.88 ; $M = 0.84$), suggesting a strong higher order or general factor. The SO factor model and the

TABLE 1: Model fit indices for confirmatory factor analysis models tested.

CFA model	χ^2	df	CFI	TLI	RMSEA	CI	SRMR	AIC
One-factor	1476.33*	135	0.760	0.727	0.112	0.107–0.118	0.077	30162.06
One-factor†	718.81*	127	0.894	0.872	0.077	0.072–0.083	0.062	28919.73
Four-factor	394.56*	129	0.952	0.943	0.051	0.045–0.057	0.038	28395.42
SO factor	399.47*	131	0.952	0.944	0.051	0.045–0.057	0.039	28401.03
Bifactor	322.57*	117	0.963	0.952	0.047	0.041–0.053	0.027	28301.38

CFA, confirmatory factor analysis; df, degrees of freedom; CFI, comparative fit index; TLI, Tucker–Lewis index; RMSEA, root mean square error of approximation; SRMR, standardised root mean square residual; AIC, Akaike information criterion; SO, second-order; CI, 90% confidence interval.

†, One-factor model with the eight freed correlated residuals (method artefacts).

*, Significant loadings ($p < 0.01$).

four-factor model showed negligible differences in model fit. The bifactor model showed a clear improvement in model fit compared with the SO factor model ($\Delta\chi^2 = 58.64$, $p < 0.01$), $\Delta CFI = 0.011$, $\Delta TLI = 0.012$, $\Delta RMSEA = 0.016$). The results showed that the proportionality constraints in the SO factor model had been violated and should be interpreted with caution. As shown in Table 2, the bifactor model had high general factor (GF) loadings (GF: $\lambda = 0.60$ – 0.71 ; $M = 0.66$) and overall weaker group-specific factor loadings ($\lambda = 0.07$ – 0.65 ; $M = 0.43$). Clearly, the GF was much better defined than the group-specific factors.

The bifactor CFA model (see Table 2) was further analysed using the appropriate factor strength indicators. The H indicator suggested that the GF was well defined ($0.94 > 0.80$) and should replicate well. None of the group-specific factors appeared well defined or replicable ($H = 0.47$ – 0.65). The ECV (0.69), ωH (0.86) for the GF and PUC (0.78) values all suggested a strong GF and an essentially unidimensional factor model. The omega coefficients (ω) for the GF and group-specific factors were all highly acceptable (GF = 0.96, $P = 0.92$, $A = 0.84$, $TM = 0.90$ and $SC = 0.84$) before partialling out GF. However, the omega coefficients hierarchical (ωS) for the group-specific factors showed very low and unreliable (< 0.50) score variances ($P = 0.30$, $A = 0.17$, $TM = 0.28$, $SC = 0.24$) after partialling out the GF's score variance (0.86). Moreover, the ECV for group-specific factors ($P = 0.12$, $A = 0.05$, $TM = 0.097$, $SC = 0.05$) was very low. The FD values showed that reliable aggregated factor scores might be calculated for GF ($0.94 > 0.90$) but not for the group-specific factors ($P = 0.82$, $A = 0.73$, $TM = 0.77$, $SC = 0.74$). Thus, little evidence existed that supported substantively relevant group-specific factors for the MACE-W2FE. The ARPB index value (0.07) showed that the absolute differences in factor scores for the GF of the bifactor BSEM model and the unmodified one-factor CFA model were negligibly small (< 0.10), suggesting that an essentially unidimensional model might be considered a plausible representation of the data.

The Jrule for Mplus analysis on the unmodified one-factor model showed a total of 16% (24/153) substantive correlated residuals exceeding an EPC of 0.10, of which only eight-item pairs (5%) exceeded an EPC of 0.10 on a 95% confidence interval (see Figure 2 for a depiction of ranked correlated residuals). Interestingly, the 24 substantive correlated residuals were all item pairs located within the

TABLE 2: Factor structure results of the confirmatory factor analysis analyses.

Item	Four-factor model					One-factor models				Bifactor model					
	P	A	TM	SC	δ	1-fact	δ	1-fact†	GF	P	A	TM	SC	δ	ARPB-I
p1	0.67*	-	-	-	0.55*	0.66*	0.56*	0.67*	0.61*	0.29*	-	-	-	0.54*	-0.10
p2	0.84*	-	-	-	0.30*	0.78*	0.39*	0.76*	0.68*	0.48*	-	-	-	0.30*	-0.10
p3	0.89*	-	-	-	0.22*	0.81*	0.35*	0.78*	0.70*	0.55*	-	-	-	0.21*	-0.11
p4	0.79*	-	-	-	0.38*	0.74*	0.45*	0.74*	0.66*	0.43*	-	-	-	0.39*	-0.13
p5	0.88*	-	-	-	0.22*	0.79*	0.37*	0.76*	0.67*	0.60*	-	-	-	0.19*	-0.13
p6	0.82*	-	-	-	0.33*	0.77*	0.40*	0.78*	0.69*	0.44*	-	-	-	0.33*	-0.13
a1	-	0.78*	-	-	0.41*	0.63*	0.60*	0.62*	0.66*	-	0.36*	-	-	0.44*	0.06
a2	-	0.83*	-	-	0.28*	0.66*	0.56*	0.66*	0.70*	-	0.65*	-	-	0.09*	0.05
a3	-	0.70*	-	-	0.49*	0.67*	0.55*	0.68*	0.71*	-	0.07*	-	-	0.49*	0.04
tm1	-	-	0.77*	-	0.40*	0.64*	0.59*	0.63*	0.60*	-	-	0.50*	-	0.39*	-0.05
tm2	-	-	0.80*	-	0.38*	0.70*	0.51*	0.69*	0.69*	-	-	0.40*	-	0.37*	-0.01
tm3	-	-	0.78*	-	0.40*	0.66*	0.56*	0.66*	0.64*	-	-	0.46*	-	0.38*	-0.02
tm4	-	-	0.78*	-	0.39*	0.69*	0.52*	0.68*	0.68*	-	-	0.38*	-	0.40*	0.00
tm5	-	-	0.77*	-	0.42*	0.66*	0.57*	0.65*	0.62*	-	-	0.47*	-	0.40*	-0.05
tm6	-	-	0.74*	-	0.46*	0.66*	0.57*	0.65*	0.64*	-	-	0.37*	-	0.46*	-0.02
sc1	-	-	-	0.79*	0.38*	0.63*	0.60*	0.62*	0.62*	-	-	-	0.50*	0.37*	0.00
sc2	-	-	-	0.71*	0.49*	0.63*	0.61*	0.63*	0.65*	-	-	-	0.27*	0.50*	0.03
sc3	-	-	-	0.88*	0.22*	0.70*	0.51*	0.69*	0.71*	-	-	-	0.54*	0.21*	0.03
Corr. resid.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
p2/p3	-	-	-	-	-	-	-	0.38*	-	-	-	-	-	-	-
p2/p5	-	-	-	-	-	-	-	0.43*	-	-	-	-	-	-	-
p3/p5	-	-	-	-	-	-	-	0.54*	-	-	-	-	-	-	-
a1/a2	-	-	-	-	-	-	-	0.48*	-	-	-	-	-	-	-
tm1/tm5	-	-	-	-	-	-	-	0.42*	-	-	-	-	-	-	-
tm2/tm4	-	-	-	-	-	-	-	0.41*	-	-	-	-	-	-	-
tm3/tm6	-	-	-	-	-	-	-	0.45*	-	-	-	-	-	-	-
sc1/sc3	-	-	-	-	-	-	-	0.50*	-	-	-	-	-	-	-
P	1.00	-	-	-	-	-	-	-	-	-	-	-	-	-	0.06 [†] ARPB
A	0.68	1.00	-	-	-	-	-	-	-	-	-	-	-	-	0.07 [†] ARPB
TM	0.67	0.75	1.00	-	-	-	-	-	-	-	-	-	-	-	-
SC	0.70	0.74	0.66	1.00	-	-	-	-	-	-	-	-	-	-	-
ω	0.92	0.82	0.90	0.84	-	0.94	-	-	0.96	0.92	0.84	0.90	0.84	-	-
SOF	0.81	0.88	0.83	0.84	-	-	-	-	-	-	-	-	-	-	-
H	-	-	-	-	-	0.94	-	-	0.94	0.65	0.47	0.58	0.45	-	-
FD	-	-	-	-	-	0.97	-	-	0.94	0.83	0.85	0.77	0.76	-	-
ω HS	-	-	-	-	-	-	-	-	0.86	0.30	0.18	0.28	0.26	-	-
ECV	-	-	-	-	-	-	-	-	0.69	0.12	0.05	0.10	0.05	-	-
PUC	-	-	-	-	-	-	-	-	77	-	-	-	-	-	-

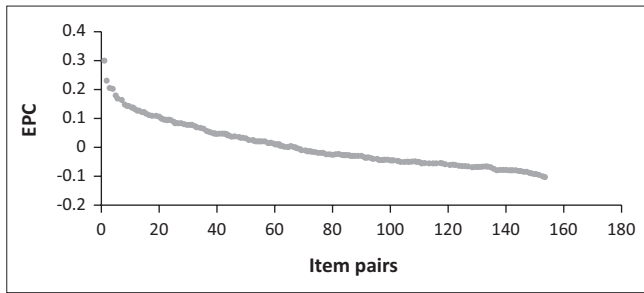
CFA, confirmatory factor analysis; P, work-family perspectives; A, work-family affect; TM, work-family time management; SC, work-family social capital; δ , residual variance; 1-fact, one-factor model; 1-fact†, one-factor model with correlated residuals; Corr. resid., correlated residuals; GF, general factor; ARPB-I, item level parameter bias; ARPB, absolute mean relative parameter bias; ω , coefficient omega; SOF, second-order factor loadings; H, construct replicability; FD, factor determinacy; ω HS, coefficient omega hierarchical; ECV, explained common variance; PUC, percentage of uncontaminated correlations.

†, One-factor model with the eight freed correlated residuals (method artefacts).

*, Significant loadings ($p < 0.01$).

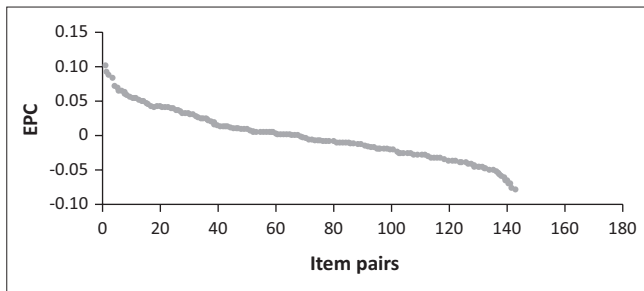
same group-specific factor of the bifactor model, pointing to shared residual variance that could substantively be explained after closer inspection. Incidentally, the eight most substantive correlated residuals were from item pairs located within each group-specific factor of the bifactor model. The eight most substantive correlated residuals (item pairs: p2/p3, p2/p5, p3/p5, a1/a2, tm1/tm5, tm2/tm4, tm3/tm6 and sc1/sc3) showed statistically significant MI values (10.42–139.99, $M = 87.45$) and notable EPC values (0.11–0.30, $M = 0.19$). Interestingly, all the correlated residual item pairs had parallel wording, similar sentence structure and semantics, and conceptually their meaning was the same, signifying item-specific method artefacts and, most likely, item redundancy (i.e. tm2: 'My family life is improved by managing my pace at work' vs. tm4: 'My family life is improved by keeping a sufficient pace at work'; sc1:

'My family life is improved by maintaining good relationships with my colleagues' vs. sc3: 'My family life is improved by having good relationships at work'). These measurement method artefacts were consequently specified as unconstrained correlated residuals in the one-factor model (see Figure 1e). The global fit indices improved significantly to obtain marginal to reasonable goodness of fit (see Table 1). The values obtained for the correlated residuals in the one-factor model were all statistically significant and had a moderate-to-large effect size ($r = 0.41$ – 0.54) (see Table 2). Values exceeding 0.20 should be regarded as noticeable and values around 0.30 as important in terms of classical test theory (Muthén & Asparouhov, 2012). The residual factors in the bifactor model can mostly be explained as item-specific method artefacts (e.g. 'Ask the same question and get the same answer'). All of the remaining correlated



EPC, expected parameter changes.

FIGURE 2: Ranked correlated residuals: One-factor model.



EPC, expected parameter changes.

FIGURE 3: Ranked correlated residuals: One-factor model with eight method artefacts freed.

residuals (i.e. 146) showed trivial misspecifications with EPC values within the range of -0.10 – 0.10 after freeing the eight misspecified item pairs with the highest correlated residuals in the model (see Figure 3 for a depiction of ranked correlated residuals). Freeing the correlated residuals for the eight most important item-specific method artefacts had a trivial effect on the factor loadings of the one-factor model (see Table 2) and a large effect on the model fit indices (see Table 1). Thus, demonstrating the sensitivity of the model fit indices for the eight most important model misspecifications, ascribed to bloated specifics in the highly restricted unidimensional model. Freeing the correlated residuals improved the model's overall factor loading bias (ARPB) from 0.07 to 0.06 (see Table 2). However, specifying method factors are preferred over correlated residuals for they explicitly estimate construct-irrelevant sources of variance where correlated residuals simply partial them out (Morin, Katrin Arens, & Marsh, 2016). The group-specific factors in bifactor model effectively represent method factors in this study. Items p2–p6's ARPB-I values varied between -0.10 and -0.13 ($M = -0.12$), causing the most factor loading bias in the unidimensional model. These items from the work–family perspectives factor shared unique variance not shared by the remaining items in the one-factor model. However, the factor strength indices showed that the unique variance was trivial and insufficient to be interpreted substantively as a distinct factor and could therefore be included as part of the model without biasing the score interpretations.

Overall, the evidence suggested that model misfit in the highly restricted one-factor CFA model could be attributed mainly to the cumulative and combined effect of trivial

TABLE 3: Approximate invariance: One-factor (unidimensional) model for genders.

Item	Males ($n = 235$)		Females ($n = 549$)	
	One-factor	AL	One-factor	AL
p1*	0.61	0.60	0.69	0.55
p2*	0.72	0.59	0.81	0.66
p3*	0.84	0.66	0.80	0.60
p4*	0.75	0.61	0.75	0.57
p5*	0.80	0.65	0.79	0.60
p6*	0.79	0.64	0.77	0.58
a1*	0.57	0.56	0.65	0.60
a2*	0.63	0.58	0.68	0.57
a3*	0.65	0.57	0.67	0.57
tm1*	0.58	0.45	0.66	0.52
tm2*	0.62	0.50	0.73	0.56
tm3*	0.59	0.59	0.69	0.52
tm4*	0.60	0.48	0.73	0.57
tm5*	0.59	0.46	0.67	0.55
tm6*	0.61	0.50	0.67	0.52
sc1 ^N *	0.74	0.67	0.59	0.45
sc2*	0.60	0.58	0.64	0.56
sc3*	0.78	0.66	0.68	0.51
Model fit statistics				
AIC	8724.27	-	21319.07	-
Chi-square*	708.98	-	1127.53	-
df	135	-	135.00	-
RMSEA	0.14	-	0.12	-
90% CI	0.125–0.144	-	0.110–0.122	-
CFI	0.69	-	0.76	-
TLI	0.65	-	0.73	-
SRMR	0.10	-	0.08	-
ω	0.93	-	0.94	-
FD	0.97	-	0.97	-

AL, loadings after alignment optimisation; p, work–family perspectives items; a, work–family affect item; tm, work–family time management item; sc, work–family social capital item; AIC, Akaike information criterion; df, degrees of freedom; RMSEA, root mean square error of approximation; CI, confidence interval; CFI, comparative fit index; TLI, Tucker–Lewis index; SRMR, standardised root mean square residual; ω , coefficient omega; FD, factor determinacy; ^N, Noninvariant.

*, Statistically significant parameter estimates ($p < 0.01$).

substantive multidimensionality, item-specific method artefacts and random noise (i.e. white noise) ascribed to imperfect indicators typically obtained in self-report questionnaire data (Asparouhov & Muthén, 2017). Thus, suggesting a plausible and parsimonious model was being rejected (i.e. type 1 error) by the goodness-of-fit indices because of large numbers of trivial model misspecifications aggravated by a large sample size.

In conclusion, adopting the more parsimonious essentially unidimensional factor model (i.e. the one with higher degrees of freedom = 135) for the MACE-W2FE instead of the more complex bifactor model (i.e. the one with lower degrees of freedom = 117) that contains an unbiased general factor can be considered justified and of practical value for applied researchers. Irrespective of showing weak global model fit, the MACE-W2FE one-factor model showed negligible bias and can be used with confidence in subsequent SEM modelling with external variables (Reise et al., 2013). The results showed that a rejection of the unidimensional one-factor CFA model based on the values of global model fit indices alone would have been unjustified.

Gender invariance testing

The one-factor models tested for invariance represented a well-identified CFA model with high factor loadings for each gender group (see Table 3). The one-factor loadings after alignment optimisation (see AL column in Table 3) was used for comparison purposes. The global model fit indices for the male group were observably lower than those for the female group, and this could be ascribed to the large difference in sample size (Kyriazos, 2018). The probability of global fit indices rejecting a non-substantively misspecified unidimensional model increases with decreasing sample sizes (Marsh et al., 2004). Having considered the likely cumulative effect of numerous trivial correlated residual misspecifications on model fit, it would be reasonable to conclude that the measurement model sufficiently represented both groups' data. The one-factor scale difference in means ($\bar{X}^{\text{Females}} - \bar{X}^{\text{Males}} = 0.027$) was negligible and statistically insignificant for gender groups. The factor loading of only one item (i.e. sc1) was flagged as being non-invariant, representing 6% of all items. With such a low level of non-invariance (< 25%), estimating group-specific factor means and variances can be expected to produce accurate results. Excluding the one item from the scale may also be considered without jeopardising scale validity for this item is in one of the item pairs showing a substantive correlated residual and, most likely, item redundancy. The omega reliability statistic (males = 0.934; females = 0.944) and the FD statistic (males = 0.971; females = 0.973) had approximately the same values for each group, showing high reliability and determinacy respectively.

Discussion

This study had three objectives within the South African sample surveyed: determine the dimensionality of the MACE-W2FE, test the scale for gender invariance and demonstrate the usefulness of extended CFA analysis techniques. This study supported hypotheses H1a in that the MACE-W2FE was essentially a unidimensional measurement model. Hypothesis H1b is supported in that the multidimensional elements of the MACE-W2FE are not distinct and substantive constructs. In addition, hypotheses H2 is supported in that the proportionality constraints of the SO factor model for the MACE-W2FE had been violated. Lastly, hypotheses H3 is supported in that the MACE-W2FE would show approximate configural, measurement and scalar invariances for gender groups.

In line with the substantive-methodology synergy framework, we discuss the substantive and methodological findings of this study in the 'Substantive findings' and the 'Methodological findings' sections, respectively. Thereafter we make concluding remarks about the study, refer to the study's limitations and make recommendations for further study.

Substantive findings

The study found that the unidimensional model best represented the general construct of WFE (Greenhaus &

Powell, 2006) and can be defined as the extent to which a variety of resources from work and family roles have the capacity to encourage individuals and to provide positive experiences, which enhance the individuals' quality of life (performance and positive affect) in the other role.

It also found that Greenhaus and Powell's (2006) work–role resources (skills, perspectives, psychological, physical and social capital) that affected the family role were reflected in the four dimensions of the MACE-W2FE. The heterogeneous content from the four dimensions was reflected as shared variance in the unidimensional model of the MACE-W2FE and enhanced the construct validity of the scale. Moreover, the evidence suggested that the MACE-W2FE reflected a broader unidimensional construct and not the distinct multidimensional constructs for which it had been developed originally. In conclusion, the data indicated that the MACE-W2FE supported an essentially unidimensional model consisting of a variety of items that reflected the variety of resources proposed in the WFE model of Greenhaus and Powell (2006).

The high intercorrelations between homogeneous item groupings of the four-factor model of the MACE-W2FE suggested the WFE construct might be hierarchical (i.e. manifesting strong common variance for group-specific factors) – a characteristic accepted almost universally as inherent to correlated multifactor psychological constructs (Clark & Watson, 1995). The hierarchical nature of item variances can be ascribed to people's responding to items at multiple conceptual levels (i.e. general and specific levels) (Rodriguez et al., 2016a). As such, WFE can be understood as a general experience directed by particular events or outcomes. Concerning the hierarchical nature of item variances, it may be relevant to note that a researcher, when trying to measure a specific domain of a general construct, faces the challenge that the diversity of the manifestations of the construct in that specific domain diminishes quickly, resulting in the researcher running out of unique questions (Rodriguez et al., 2016a). Therefore, the researcher may include questions that differ little in content. Such subdomain item redundancy has been termed 'bloated specifics' (Cattell, 1978). This study showed that the group-specific factors in the four-factor model of the MACE-W2FE contained little substance after the common variance in the general factor had been partialled out. Such factors, which are (arti)factors with little substance, are common occurrences in published scales (Rodriguez et al., 2016a).

Previous criterion-related validation studies on the MACE also provided support for the plausibility of the MACE-W2FE consisting of an essentially unidimensional construct as opposed to four distinct constructs (De Klerk et al., 2015; Koekemoer et al., 2017; Marais et al., 2014). It was found that the group-specific four-factor model showed moderate correlates (i.e. $r = 0.26-0.43$) with measures of related constructs (i.e. job satisfaction and other WFE outcomes) (De Klerk et al., 2015). Consequently, researchers may

argue that the subscales show differential correlates with related constructs and that they, therefore, show construct uniqueness. This contention is not accurate as any two variables that are not perfectly correlated will show differential correlates with a third variable as each is a mixture of the same general factor and a distinct group-specific factor (Rodriguez et al., 2016a). The current study indicated that the group-specific factors of the MACE-W2FE showed little construct uniqueness and that any correlates with a third variable could be attributed to the underlying general factor. It would be reasonable to accept that the general factor in a higher order model depicts high levels of common variance shared by all the items in the group-specific factors and therefore shows high criterion-related correlates. Koekemoer et al. (2017) and Marais et al. (2014) supported this notion by showing that the correlations (i.e. $r = 0.50-0.66$) between the general factor of the SO factor model of the MACE-W2FE and the third variables (i.e. job satisfaction and other WFE outcomes) were much higher than correlations obtained for the group-specific four-factor model (i.e. $r = 0.26-0.43$) (De Klerk et al., 2015).

Thus, it would be reasonable to suggest that the essentially unidimensional model of the MACE-W2FE with its underlying general factor is, when compared with the group-specific four-factor model, a more robust representation of Greenhaus and Powell's (2006) conceptualisation of WFE theory.

The MACE-W2FE unidimensional measurement model clearly reflects Rodriguez et al.'s (2016a) notion that the social sciences can be best served by positing a strong theory for a general construct and having a thorough understanding of the construct and its links to the processes of item responses, thereby ensuring it is measured well.

The current study corroborated the findings of De Klerk et al. (2013) that gender groups were comparable on the MACE-W2FE and showed similar scores. Yet, Van Steenberghe et al. (2007) found that women experienced more WFE than men, whereas Rothbard (2001) found the opposite. It is clear that more studies are needed to obtain clarity about gender differences regarding WFE.

Methodological findings

The bifactor modelling, local indicator misfit analyses and approximate invariance testing proved to be useful tools for understanding the sources of item variances and the psychometric functioning of the proposed multidimensional or unidimensional model of the MACE-W2FE (Rodriguez et al., 2016a). Our study showed how the factor strength indices used in combination with bifactor analyses and local indicator analyses successfully resolved the dimensionality issues of the MACE-W2FE, whereas global CFA fit indices showed limited value in this regard. The data supported an essentially unidimensional model for the

MACE-W2FE, although there was a minor element of multidimensionality. Our finding supported the finding of Rodriguez et al. (2016a) that the scores for the 50 measures of the unidimensional (reportedly multidimensional) models they studied were highly resilient to the biasing effects of multidimensionality. It is conceivable that researchers reject an essentially unidimensional model based on model fit indices alone because it contains a mixture of trivial multiple dimensional substantive elements, method artefacts and white noise, which, according to common belief, cause such a model to defy meaningful interpretation. However, Rodriguez et al. (2016a) alluded to the work of Gustafsson and Alberg-Bengtsson (2010) in stating that:

[I]t is a myth [that essentially unidimensional models defy meaningful interpretation]: when correlated items are aggregated together, and they all share a single common factor, that the more items that are grouped, the more the total score reflects that common latent variable, regardless of the dimensionality (p. 232).

Cronbach (1951) knew this – he demonstrated this principle in his original coefficient alpha paper, which has been widely cited. In addition, Bentler (2009) stated that global fit indices were unlikely to show good model fit for unidimensional CFA models where the number of items was large. A CFA unidimensional model has large degrees of freedom and may be considered a highly restrictive model, but, when compared with an alternative model (i.e. a bifactor model) with lower degrees of freedom, it is the more parsimonious model.

We further showed how thoroughly considering local misspecification information could assist in adjudicating model fit. More specifically, we found that the accumulative effect of trivial correlated residual misspecifications could explain the misfit on the global fit indices for the one-factor model. Moreover, the statistical power of the correlated residual misspecifications was all acceptable (> 0.8), making type 1 error and the need for a verification study sample an issue of lesser concern. After considering all the information, we could make the reasonable conclusion that the one-factor model represented the data reasonably well. This finding is consistent with arguments against the simplistic conceptualisation of the dimensionality of psychological data and the value of global fit indices as a sole means of adjudicating model (mis)fit (Rodriguez et al., 2016a) arguments that deemed especially relevant in the case of highly restrictive unidimensional models consisting of numerous items.

Moreover, evidence was compelling that the eight most important item residual correlates in the one-factor model were method artefacts reflected in the residual factors of the bifactor model and contributed to the four-factor model of the MACE-W2FE being pseudo-specific and deceptive. Some researchers may argue that a good global model fit can be

obtained by reducing the number of items in the one-factor model. However, it would be counter-productive to shorten a measure of a broadly defined construct such as WFE simply to comply with goodness-of-fit indices' cut-off criterion (Marsh et al., 2004). This would surely jeopardise the coverage of all the subdomains of importance in a general construct such as WFE.

In addition, we provided strong methodological arguments and empirical evidence that the violation of proportionality constraints and the related challenges associated with score interpretation could make the use of SO factor models of the MACE-W2FE in particular (Koekemoer et al., 2017; Marais et al., 2014) and the WFE in general (Rastogi et al., 2018; Russo et al., 2018) less ideal.

The approximate invariance test technique proved helpful in making valid comparisons between gender groups without having to make questionable model modifications to obtain exact measurement invariance or seek partial invariance, which can be a cumbersome process (Asparouhov & Muthén, 2014).

Practical recommendations

This study showed that an essentially unidimensional measurement model of the MACE-W2FE should be included in further studies on WFE with external variables. However, the essentially unidimensional measurement model's low goodness of fit indices may adversely reflect on SEM models overall model fit. However, the factor strength indicators showed the unidimensional model can be incorporated as an aggregated score in SEM models with negligible biasing effects on regression paths or a reduction in measurement precision. Researchers may also consider forming item parcels through collapsing highly correlated item pairs or triplets from similar content subdomains so as to simplify the one-factor model for use in subsequent SEM analyses (Rodriguez et al., 2016a). Where model complexity and convergence are not an issue in an SEM model with external variables, researchers may consider including the bifactor measurement model and treat the group-specific factors as method factors.

Study limitations

An important limitation of the study is that a convenience sample was used and that the participants were limited to employees in the South African work environment. Therefore, sample homogeneity was promoted at the cost of external validity. A larger and randomly selected sample stretching across nationalities, industries, job types, work conditions and cultures would have better served the purposes of the study.

Furthermore, confirming the dimensionality and gender invariance of the MACE-W2FE does not render it a valid measure of WFE. The MACE-W2FE items may need

reviewing for redundancy, and ongoing construct- and criterion-related research will be beneficial for the future use of the measure.

Conclusion

In this study, we thoroughly investigated the MACE-W2FE at different levels of analysis and used various statistical indicators. The rigor of analyses enabled us to make an informed choice about a robust MACE-W2FE measurement model that best reflected the WFE theory.

With this study, we hoped to inspire applied researchers in South Africa to pursue a 'substantive-methodology synergy' approach by utilising advanced statistical tools with the power and flexibility to facilitate an in-depth and thorough analysis of hypothesised measurement models. Such rigor in scientific endeavour can only benefit the quality of the quantitative measures used for research in the management sciences.

Work-family enrichment research is on the increase because WFE has been shown to not only improve people's quality of life but also enhance work engagement, job satisfaction, work vigour, job dedication and general career satisfaction, which all contribute to human performance (De Klerk et al., 2013, 2015; Marais et al., 2014; Van Steenbergen et al., 2007). The need for a robust WFE measure backed by strong theory that will allow further studies to be conducted in the field has been well articulated (De Klerk et al., 2013). Finally, the MACE-W2FE appears to be gender invariant, which opens up opportunities for further research on gender differences in the domain of WFE in the future world of work.

Acknowledgements

Competing interests

The authors have declared that no competing interests exist.

Authors' contributions

All authors contributed equally to this work.

Funding information

This work is based on research supported, in part, by the National Research Foundation of South Africa (Grant Number 103796).

Data availability statement

The data that support the findings of this study are available from the corresponding author, Pieter Schaap, upon reasonable request.

Disclaimer

The views and opinions expressed in this article are those of the authors and do not necessarily reflect the official policy or position of any affiliated agency of the authors.

References

- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling*, 21(4), 495–508. <https://doi.org/10.1080/10705511.2014.919210>
- Asparouhov, T., & Muthén, B. (2017). *Prior-posterior predictive p-values*. Retrieved from <https://www.statmodel.com/download/PPPP.pdf>
- Bentler, P.M. (2009). Alpha, dimension-free, and model-based internal consistency reliability. *Psychometrika*, 74(1), 137–143. <https://doi.org/10.1007/s11336-008-9100-1>
- Carlson, D.S., Grzywacz, J.G., & Zivnuska, S. (2009). Is work-family balance more than conflict and enrichment? *Human Relations*, 62(10), 1459–1486. <https://doi.org/10.1177/0018726709336500>
- Carlson, D.S., Kacmar, K.M., Wayne, J.H., & Grzywacz, J.G. (2006). Measuring the positive side of the work-family interface: Development and validation of a work-family enrichment scale. *Journal of Vocational Behavior*, 68(1), 131–164. <https://doi.org/10.1016/j.jvb.2005.02.002>
- Cattell, R.B., & Tsujioka, B. (1964). The importance of factor-trueness and validity, versus homogeneity and orthogonality, in test scales. *Educational and Psychological Measurement*, 24(1), 3–30. <https://doi.org/doi:10.1177/00131646402400101>
- Cattell, R.B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York, NY: Springer.
- Cerny, B.A., & Kaiser, H.F. (1977). A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behavioral Research*, 12(1), 43–47. https://doi.org/10.1207/s15327906mbr1201_3
- 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>
- Chen, F.F., West, S.G., & Sousa, K.H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5
- Clark, L.A., & Watson, D. (1995). Construct validity basic issue in objective scale development. *Psychological Assessment*, 7(3), 309–319. <https://doi.org/10.1037/1040-3590.7.3.309>
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155–159. <https://doi.org/10.1037/0033-2909.112.1.155>
- Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
- De Bondt, N., & Van Petegem, P. (2015). Psychometric evaluation of the overexcitability questionnaire—two applying Bayesian structural equation modeling (BSEM) and multiple-group BSEM-based alignment with approximate measurement invariance. *Frontiers in Psychology*, 6, 1–17. <https://doi.org/10.3389/fpsyg.2015.01963>
- De Klerk, M., Nel, J.A., Hill, C., & Koekemoer, E. (2013). The development of the MACE work-family enrichment instrument. *SA Journal of Industrial Psychology*, 39(2), 1–16. <https://doi.org/10.4102/sajip.v39i2.1147>
- De Klerk, M., Nel, J.A., & Koekemoer, E. (2015). Work-to-family enrichment: Influences of work resources, work engagement and satisfaction among employees within the South African context. *Journal of Psychology in Africa*, 25(6), 537–546. <https://doi.org/10.1080/14330237.2015.1124606>
- DeVellis, R.F. (2003). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage.
- Dubihlela, J., & Dhurup, M. (2013). Negative work-family and family-work conflicts and the relationship with career satisfaction among sport coaching officials. *African Journal for Physical Health Education, Recreation and Dance*, 19(Suppl 2), 177–192.
- Eby, L.T., Casper, W.J., Lockwood, A., Bordeaux, C., & Brinley, A. (2005). Work and family research in IO/OB: Content analysis and review of the literature (1980–2002). *Journal of Vocational Behavior*, 66(1), 124–197. <https://doi.org/10.1016/j.jvb.2003.11.003>
- Frone, M.R., & Yardley, J.K. (1997). Developing and testing an integrative model of the work-family interface. *Journal of Vocational Behavior*, 167(50), 145–167. <https://doi.org/10.1006/jvbe.1996.1577>
- Garrido, C.C., González, D.N., Seva, U.L., & Piera, P.J.F. (2019). Multidimensional or essentially unidimensional? A multi-faceted factoranalytic approach for assessing the dimensionality of tests and items. *Psicothema*, 31(4), 450–457. <https://doi.org/10.7334/psicothema2019.153>
- Gignac, G.E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence*, 55, 57–68. <https://doi.org/10.1016/j.intell.2016.01.006>
- Greenhaus, J.H., & Beutell, N.J. (1985). Sources of conflict between work and family roles. *Academy of Management Review*, 10(1), 76–86. <https://doi.org/10.2307/258214>
- Greenhaus, J.H., & Powell, G.N. (2006). When work and family are allies: A theory of work-family enrichment. *Academy of Management Review*, 31(1), 77–92. <https://doi.org/10.5465/AMR.2006.19379625>
- Greiff, S., & Heene, M. (2017). Why psychological assessment needs to start worrying about model fit. *European Journal of Psychological Assessment*, 33(5), 313–317. <https://doi.org/10.1027/1015-5759/a000450>
- Gustafsson, J.-E., & Alberg-Bengtsson, L. (2010). Unidimensionality and interpretability of psychological instruments. In S.E. Embretson (Ed.), *Measuring psychological constructs: Advances in model-based approaches* (pp. 97–121), Washington, DC: American Psychological Association.
- Hanson, G.C., Hammer, L.B., & Colton, C.L. (2006). Development and validation of a multidimensional scale of perceived work-family positive spillover. *Journal of Occupational Health Psychology*, 11(3), 249–265. <https://doi.org/10.1037/1076-8998.11.3.249>
- Hayduk, L.A. (2014). Seeing perfectly fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse. *Educational and Psychological Measurement*, 74(6), 905–926. <https://doi.org/10.1177/0013164414527449>
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 37–41. <https://doi.org/10.1080/10705519909540118>
- Jaga, A., Bagraim, J., & Williams, Z. (2013). Workfamily enrichment and psychological health. *SA Journal of Industrial Psychology*, 39(2), 1–10. <https://doi.org/10.4102/sajip.v39i2.1143>
- Jiang, H., & Men, R.L. (2017). Creating an engaged workforce: The impact of authentic leadership, transparent organizational communication, and work-life enrichment. *Communication Research*, 44(2), 225–243. <https://doi.org/10.1177/0093650215613137>
- Kacmar, K.M., Crawford, W.S., Carlson, D.S., Ferguson, M., & Whitten, D. (2014). A short and valid measure of work-family enrichment. *Journal of Occupational Health Psychology*, 19(1), 32–45. <https://doi.org/10.1037/a0035123>
- Koekemoer, E., Mostert, K., & Rothmann, I., Jr. (2010). Interference between work and nonwork roles: The development of a new South African instrument. *SA Journal of Industrial Psychology*, 36(1), 1–15. <https://doi.org/10.4102/sajip.v36i1.907>
- Koekemoer, E., Olckers, C., & Nel, C. (2020). Work-family enrichment, job satisfaction, and work engagement: The mediating role of subjective career success. *Australian Journal of Psychology*, 72(4), 1–12. <https://doi.org/10.1111/ajpy.12290>
- Koekemoer, E., Strasheim, A., & Cross, R. (2017). The influence of simultaneous interference and enrichment in work-family interaction on work-related outcomes. *South African Journal of Psychology*, 47(3), 330–343. <https://doi.org/10.1177/0081246316682631>
- Kyriazos, T.A. (2018). Applied psychometrics: Sample size and sample power considerations in factor analysis (EFA, CFA) and SEM in general. *Psychology*, 9(8), 2207–2230. <https://doi.org/10.4236/psych.2018.98126>
- Marais, E., De Klerk, M., Nel, J.A., & De Beer, L. (2014). The antecedents and outcomes of work-family enrichment amongst female workers. *SA Journal of Industrial Psychology*, 40(1), Art. #1186, 14 pages. <https://doi.org/10.4102/sajip.v40i1.1186>
- Marsh, H.W., & Hau, K-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, 32(1), 151–170. <https://doi.org/10.1016/j.cedpsych.2006.10.008>
- Marsh, H.W., Hau, K-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: A Multidisciplinary Journal*, 11(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2
- McDonald, R.P. (1999). *Test Theory: A Unified Treatment*. New York, NY: Routledge Taylor & Francis Group.
- McNeish, D., An, J., & Hancock, G.R. (2018). The thorny relation between measurement quality and fit index cut-offs in latent variable models. *Journal of Personality Assessment*, 100(1), 43–52. <https://doi.org/10.1080/00223891.2017.1281286>
- Morin, A.J.S., Katrin Arens, A., & Marsh, H.W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*, 23(1), 116–139. <https://doi.org/10.1080/10705511.2014.961800>
- Mostert, K. (2011). Job characteristics, work-home interference and burnout: Testing a structural model in the South African context. *International Journal of Human Resource Management*, 22(5), 1036–1053. <https://doi.org/10.1080/09585192.2011.556777>
- Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods*, 17(3), 313–335. <https://doi.org/10.1037/a0026802>
- Muthén, B., & Asparouhov, T. (2018). Recent methods for the study of measurement invariance with many groups: Alignment and random effects. *Sociological Methods and Research*, 47(4), 637–664. <https://doi.org/10.1177/0049124117701488>
- Muthén, L.K., & Muthén, B. (2017). *Mplus 8 user's guide*. Los Angeles, CA: Muthén & Muthén.
- Oberski, D.L. (2009). *Jrule for Mplus*. Tilburg: Tilburg University.
- Opie, T.J., & Henn, C.M. (2013). Work-family conflict and work engagement among mothers: Conscientiousness and neuroticism as moderators. *SA Journal of Industrial Psychology*, 39(1), 1–12. <https://doi.org/10.4102/sajip.v39i1.1082>
- Rastogi, M., Karatepe, O.M., & Mehmetoglu, M. (2018). Linking resources to career satisfaction through work-family enrichment. *Service Industries Journal*, 39(11–12), 855–876. <https://doi.org/10.1080/02642069.2018.1449835>
- Reise, S.P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>
- Reise, S.P., Bonifay, W.E., & Haviland, M.G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, 95(2), 129–140. <https://doi.org/10.1080/00223891.2012.725437>
- Reise, S.P., Moore, T.M., & Haviland, M.G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544–559. <https://doi.org/10.1080/00223891.2010.496477>

- Reise, S.P., Scheines, R., Widaman, K.F., & Haviland, M.G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5–26. <https://doi.org/10.1177/0013164412449831>
- Rodríguez, A., Reise, S.P., & Haviland, M.G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*(3), 223–237. <https://doi.org/10.1080/00223891.2015.1089249>
- Rodríguez, A., Reise, S.P., & Haviland, M.G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150. <https://doi.org/10.1037/met0000045>
- Ropovik, I. (2015). A cautionary note on testing latent variable models. *Frontiers in Psychology, 6*, 1–8. <https://doi.org/10.3389/fpsyg.2015.01715>
- Rothbard, N.P. (2001). Enriching or depleting? The dynamics of engagement in work and family roles. *Administrative Science Quarterly, 46*(4), 655. <https://doi.org/10.2307/3094827>
- Russo, M., Buonocore, F., Carmeli, A., & Guo, L. (2018). When family supportive supervisors meet employees' need for caring: Implications for work-family enrichment and thriving. *Journal of Management, 44*(4), 1678–1702. <https://doi.org/10.1177/0149206315618013>
- Saris, W.E., Satorra, A., & Van Der Veld, W.M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal, 16*(4), 561–582. <https://doi.org/10.1080/10705510903203433>
- Sass, D.A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29*(4), 347–363. <https://doi.org/10.1177/0734282911406661>
- Schmitt, T.A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment, 29*(4), 304–321. <https://doi.org/10.1177/0734282911406653>
- Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment, 31*(12), 1428–1441. <https://doi.org/10.1037/pas0000623>
- Shockley, K.M., & Singla, N. (2011). Reconsidering work-family interactions and satisfaction: A meta-analysis. *Journal of Management, 37*(3), 861–886. <https://doi.org/10.1177/0149206310394864>
- Sieber, S.D. (1974). Toward a theory of role accumulation. *American Sociological Review, 39*(4), 567–578. <https://doi.org/10.2307/2094422>
- Siu, O.L., Bakker, A.B., Brough, P., Lu, C.Q., Wang, H., Kalliath, T., ... Timms, C. (2015). A three-wave study of antecedents of work-family enrichment: The roles of social resources and affect. *Stress and Health, 31*(4), 306–314. <https://doi.org/10.1002/smi.2556>
- Strauss, M.E., & Smith, G.T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology, 27*(5), 1–25. <https://doi.org/10.1146/annurev.clinpsy.032408.153639.Construct>
- Ten Brummelhuis, L.L., & Bakker, A.B. (2012). A resource perspective on the work-home interface: The work-home resources model. *American Psychologist, 67*(7), 545–556. <https://doi.org/10.1037/a0027974>
- Timms, C., Brough, P., O'Driscoll, M., Kalliath, T., Siu, O.L., Sit, C., & Lo, D. (2015). Positive pathways to engaging workers: Work-family enrichment as a predictor of work engagement. *Asia Pacific Journal of Human Resources, 53*(4), 490–510. <https://doi.org/10.1111/1744-7941.12066>
- Van Steenbergen, E.F., Ellemers, N., & Mooijart, A. (2007). How work and family can facilitate each other: Distinct types of work-family facilitation and outcomes for women and men. *Journal of Occupational Health Psychology, 12*(3), 279–300. <https://doi.org/10.1037/1076-8998.12.3.279>
- Van Zyl, P. (2020). *The development and empirical evaluation of a structural model of enrichment among female academics*. Stellenbosch University. Retrieved from <http://scholar.sun.ac.za/handle/10019.1/108212>
- Voydanoff, P. (2002). Linkages between the work-family interface and work, family and individual outcomes: An integrative model. *Journal of Family Issues, 23*(1), 138–164. <https://doi.org/10.1177/0192513x02023001007>
- Wayne, J.H., Grzywacz, J.G., Carlson, D.S., & Kacmar, K.M. (2007). Work-family facilitation: A theoretical explanation and model of primary antecedents and consequences. *Human Resource Management Review, 17*(1), 63–76. <https://doi.org/10.1016/j.hrmr.2007.01.002>
- Yung, Y.F., Thissen, D., & McLeod, L.D. (1999). On the relationship between the higher-order factor model and the hierarchical factor model. *Psychometrika, 64*(2), 113–128. <https://doi.org/10.1007/BF02294531>
- Zhang, Y., Xu, S., Jin, J., & Ford, M.T. (2018). The within and cross domain effects of work-family enrichment: A meta-analysis. *Journal of Vocational Behavior, 104*, 210–227. <https://doi.org/10.1016/j.jvb.2017.11.003>