Agustin Rahmawati¹ ; Suryanto²; Nurul Hartini³

¹,²,³Psychology Doctoral Program, Faculty of Psychology, Airlangga University

Received: March 26, 2022; Accepted: August 08, 2022; Published: September 30, 2022

Abstract
The need for the availability of work-family balance measuring instruments in the Indonesian context is becoming an increasingly urgent thing, along with the increasing interest in research on the field of work and family. This study aims to test adapting and validating the Work-Family Balance Scale developed by Carlson, et al., (2009) into the Indonesian context. The trial was conducted on 104 women who worked as policewoman. A series of adaptation processes have been carried out, ranging from translation, equivalence testing, validation, and reliability testing. Validity of content was tested through the Content Validity Index (CVI) by involving several expert reviewers, and the validity of constructs, using Confirmatory Factor Analysis (CFA). Based on this trial process, it is known that the WFB construct meets all Goodness of Fit criteria (RMSEA ≤ 0.08; SRMR ≤ 0.05; NFI ≥ 0.9; CFI ≥ 0.9). In addition, Construct Reliability (CR) values > 0.87 and Average Variance Extracted (AVE) > 0.53. Thus, the WFB Scale can be used as a valid and reliable measuring tool to measure work-family balance in the Indonesian context.

Keywords
adaptation; validation; work-family balance; scale

INTRODUCTION
Research on work-family balance has increased very sharply in recent years (Ferguson, et. al., 2012). Empirical evidence suggests that work-family balance is associated with job satisfaction and family satisfaction, organizational commitment, and family performance (Pattusamy & Jacob, 2017; Wiens, et.al., 2022). Another study found that work-family balance was associated with overall life satisfaction, mental health, and marital quality (Matysiak, et.al., 2016; Wayne et.al., 2017; Wayne, et.al., 2020). Therefore, further research on work-family balance becomes important for some of the above reasons, including in the situation of the covid-19 pandemic (Adisa, et.al., 2021).

Interest in the study of work-family balance began in the 1970s (O’Driscoll et.al. 2006) to describe the balance between individual work and personal life. For some authors, the idea of balancing different roles that ultimately result in a satisfactory balance is a challenge in itself (Landolfi & Presti, 2020).

At first, work-family balance only refers to the concept of the absence of work-family conflict (Hill, et.al., 2001). Work-family conflict is defined as a conflict between roles, where the pressures of work and fa-
Family roles conflict with each other in several ways (Greenhaus & Beutell, 1985). While Frone, et al. (2003) found that work-family conflict implies a two-way conflict; work-to-family conflict and family-to-work conflict. That the work domain interferes with the family, instead the family domain interferes with work.

With the development of research lately, the absence of conflict alone is not enough to indicate a work-family balance (Grzywacz & Carlson, 2007; Valcour, 2007; Chang, et al., 2017). The concept of work-family facilitation is the second component that needs to be considered in indicating the balance of family work. (Grzywacz & Marks, 2000; Kirchmeyer, in Grzywacz & Carlson, 2007). Therefore, some researchers use four dimensions to explain the emergence of work-family balance, including work-to-family conflict, family-to-work conflict, family-to-work facilitation, and family-to-work facilitation (Grzywacz & Marks, 2000). As with work-family conflicts, work-family facilitation also has a two-way dimension; the domain of work can be a family facilitation, while the family domain can be a facilitation for work affairs (Frone, 2003).

On the way, progressively researchers began to recognize work-family balance as a different construct (Grzywacz & Carlson, 2007; Greenhaus & Allen, 2011). Instead of considering work-family balance as a reflection of the four dimensions above, the researchers found that work-family balance generally describes a thorough evaluation of compatibility between work roles and family roles. Some authors try to distinguish conceptually and empirically support the difference between work-family conflict, work-family facilitation/enrichment, and work-family balance (Carlson et al., 2009).

Several different definitions of work-family balance were put forward. Greenhaus, et al. (2003) say that work-family balance is associated with the extent to which individuals feel involved and feel satisfaction, both in the field of work and family. Hirsch, et al. (2019) proposed a theoretical model that explains how people can jointly achieve work and family goals using the four resource utilization strategies. Greenhaus & Allen (2011) defines work-family balance as the extent to which an individual’s effectiveness and satisfaction in carrying out work and family roles, according to individual life priorities. Meanwhile, Wayne, et al. (2017) proposed a framework that elaborates on four conceptualizations of work-family balance: additive spillover, multiplicative, balance satisfaction, and balance effectiveness. Another concept, Grzywacz and Carlson (2007) defines the work-family balance as a thorough evaluation of the achievement of the role performed by individuals in connection with the hope of negotiation and sharing roles with partners in carrying out their roles in both the work and family domains. This understanding refers to the existence of a social dimension in realizing its responsibilities related to the role of work and family (Landolfi & Presti, 2020).

Despite all that, the work-family balance is still a relatively new concept, and many problems remain unsolved both methodologically and theoretically. This is a challenge in itself because it limits our understanding of this phenomenon. Work-family balance becomes important in addressing the problem of conflict between work and family, the level of which has increased in recent years (Morganson, et al., 2014) cause work-family balance will increase a person’s life satisfaction (Jensen, et al., 2017) and be negatively correlated with anxiety and depression (Haar, et al. 2014).

Considering the importance of pursuing a work-family balance, the researchers sought to create a measuring instrument that could identify work-family balance. With the availability of work-family balance, measuring instruments will allow organizations and families to support individuals to find balance, which will later have an impact not only on the well-being of individuals and families, but also on the interests of organizations (Landolfi & Presti, 2020; Bianchi & Milkie, 2010; Frone, 2003; Handayani, 2017; Jensen, et al., 2017; Lyu, et al., 2019).

In literature studies, independent work-family balance measuring instruments are still not widely found, because
some experts define work-family balance as the absence of work-family conflict and high levels of enrichment, so some measurements of work-family balance are based on low levels of conflict and high levels of enrichment (Fronen, 2003), including Fisher-McAuley et al. (2003) and Huffman et al. (2004) which developed a measurement of work-family balance with reference to high facilitation and low conflict. Other measuring instruments used to uncover work-family balance, for example those compiled by Milkie and Petola (1999) with single item: “How successful do you feel in balancing your paid work and family life?” and White (1999) emphasizes more on satisfaction with balance as well with single items: “Are you satisfied or dissatisfied with the balance between your job or main activity and family and home life? “, and Marks and MacDermid (1996) developed an 8-item scale of role balance.

The scale in this paper is the one proposed by Carlson et al. (2009). It is slightly different because the underlying construct focuses more on the social domain (Landolfi & Presti, 2020). This measuring instrument has been used in several previous studies and produces satisfactory reliability coefficients, including in China, Germany, and Egypt, all of which produced a reliability coefficient above 0.80 (Landolfi & Presti, 2020). The study by Omran (2016) in Egypt found a Cronbach’s alpha of .93, while Krisor et al. (2015), in Germany found a Cronbach’s alpha of .86, Lyu, et al., (2019) in China with a Cronbach’s alpha of .95.

In Indonesia, measuring instruments about work-family balance are still developing a lot. Some work-family balance researchers in the Indonesian context have not provided information related to the adaptation and validation of measuring instruments formally (Handayani, et al., 2017). Therefore, it becomes a necessity to adapt and validate the work-family balance scale from Carlson, et al. (2009) in the Indonesian context, so that it will be easier for future researchers to get a measuring tool for work-family balance as needed.

**METHOD**

WFBS was developed by Carlson et al. (2009). This measuring instrument is unidimensional and consists of 6 items, with a small number of items intended to avoid saturation of the subject because it has to respond to many measuring instruments. In his research, Carlson et al. (2009) reported an alpha cronbach reliability coefficient of .93 and all items had a loading factor above 0.77, meaning a high loading factor. Indicators with high loading factors have a higher contribution to explain their latent constructs. Conversely, indicators with low loading factors have a weak contribution to explain the latent construct. Thus, this measuring instrument has reliability and validity that is quite satisfactory.

Subjects’ responses used the Likert model, with moving ratings of 1 (strongly disagree), 2 (disagree), 3 (undecided), 4 (agree), and 5 (strongly agree). The work-family balance level is represented by the total score of the WFB scale, the higher the total score the higher the work-family balance.

In the adaptation process, researchers used the International Test Commission (ITC) Guidelines for Test Adaptation (2017) as guidelines. At the preparatory stage, researchers ask permission from the measuring instrument maker by sending a message via email. In addition, researchers also select a group of translators and experts who will be involved in the adaptation process. The selection was made based on several considerations, such as educational background, Indonesian and English language skills, and research focus.

Translation is the next process. This process is carried out by two translators, both of whom do not know each other. Translators are selected people who are experts in English and Indonesian to avoid misinterpretation. Translators are provided with information about the purpose of the research, the operational definition of each variable and the plan of the research sample so that the translator can better understand the intent and the translation purpose. At this stage, the original Scale in English is
translated into Indonesian, also called forward translation. Forward translation results are then discussed to obtain a single Indonesian translation scale called synthesis forward translation. The next translation stage is backward translation. Backward translation is the process of translating forward translation into English. Backward translation is done to ensure that the translation results in Indonesian do not deviate from the original scale. The backward translation process is carried out by two translators who do not know each other. These two translators did not see the original scale, only given the synthesis of forward translation.

The results of this backward translation are given to three expert reviewers to see the comparison of translation results with the original measuring instrument by filling in the form that has been provided. The assessment form uses rating scales as recommended in the ITC Guidelines for Adaptation (2017), with a rating range ranging from 1 to 7. Comparability indicates the level of similarity of language, phrases, terms, words, and sentences formally. Statement items that are very identical and have no difference are given a score of 1, while items that are not at all identical are given a score of 7. Similarity is the degree of similarity of meaning between two versions of statement items, although the terms used are different. Statement items that have identical meanings are scored 1, while those with very different meanings are scored 7. Sperber (2004) explains that ideal conditions occur when statement items have similarities, both in the form of language and meaning, but the similarity of meaning takes precedence over the similarity of forms. From the results of the assessment of rating scale comparability and similarity from expert reviewers, mean score calculations were carried out.

The validation process of this measuring instrument is analyzed based on the validity of the contents and the validity of the construct. The validity of the content is evidence that indicates the extent to which the contents of the test are in accordance with the intended purpose, the evidence is to establish that the test items represent the measuring region. The validity of the content on the scales used in this study was carried out using the content validity index (CVI). Polit et al (2007) explained, there are two types of CVI, namely Item-CVI (I-CVI) and Scale-CVI (S-CVI). I-CVI involves more individual validity of item content whereas S-CVI is used to measure the validity of the content of the scale as a whole. The calculation of I-CVI is done by asking for the help of expert reviewers to provide an assessment in the form of an assessment rating of 1 to 4 related to three things, namely relevancy, importance and clarity. Relevance is the extent to which the relevance of an item to the construct is measured. Importance is how important the item is when it is associated with the construction and context of the research. Clarity focuses on whether the item is clear enough and understandable. Polit et al (2007) explained the I-CVI rating from a rating of 1 - 4. Good item is the one rated 3 and 4, and the bad item is the one rated 1 and 2. Based on the procedures of Polit et al (2007) then the assessment of the validity of the contents in this research scale was made a dichotomy score, namely a score of 1 (for item rated 3 and 4) and a score of 0 (for item rated 1 and 2).

The second stage of validation is by testing measuring instruments. A total of 104 women who worked as policewomen, were married, and had children under the age of 18 were involved in this trial process. Measuring instrument trials are carried out to obtain evidence related to validity and reliability. Validity in this study is referred to as construct validity which is a type of validity that indicates the extent to which the measuring instrument reveals the theoretical construct measured. The validity of the construct is carried out using the Confirmatory Factor Analysis (CFA) of Lisrel 9.10 program to test the internal structure of the measuring instrument. In this study, the First Order Confirmatory Factor Analysis will be carried out because the work-family balance construct is a unidimensional construct. Confirmatory Factor Analysis or CFA is a statistical method used to confirm a
variable (item, indicator, observed variable, dimension) whether it is part of a construct variable or dimension (unobserved variable, latent variable) or how much observed variables are able to measure or explain unobserved variables, which are characterized by a test of validity and reliability (reliability). The CFA analysis method is used to test the questionnaire instrument, whether it is valid and reliable. The purpose of the CFA in the questionnaire test is to confirm or test whether the items questions in the questionnaire are validly explaining the construct/factor and whether the construct/factor is reliable.

The CFA test process is the first to test conformity on a measurement model with a goodness of fits value. The fit indexes used as guidelines in this study are Root Mean Square Error Approximation (RMSEA), Standardized Root Mean Square Residual (SRMR), Normed Fit Indices (NFI) and Comparative Fit Indecies (CFI). If the model is not fit, modifications are made in accordance with the advice given by the Lisrel 9.10 program which can be seen in the modification indices. After checking the suitability of the measurement model and obtaining a fit CFA model, the next step is to check the validity of the construct. Construct validity can be done in two ways, namely convergent validity testing and discriminant validity. Convergent Validity Test is a construct validity test by looking at the loading factor value of each item. If it has a loading factor value greater than 0.6 then the item is valid in a convergent measure of its dimensions (construct). As Hair, et al., (2010) and Ghozali (2014), that the reference value of loading factors of 0.50 or more is considered to have strong validation to explain their dimensions (construct). The next construct validity test is the discriminant validity test, namely by comparing the root value of the Average Variance Extracted (AVE) of each construct variable with the correlation value between the construct variables. If the root value of AVE is greater then it is concluded to have good descriptor validity.

After testing the validity of the construct, reliability testing is carried out. Reliability is the extent to which the results of a measurement process can be trusted. Measuring instrument reliability testing is carried out through measurement of construct reliability values (Construct Reliability or Composite Reliability; CR) and the average value of extracted variance (Average Variance Extracted; AVE). The construct reliability coefficient emphasizes how far the measuring indicator reflects the latent factors compiled. The greater the indicator reflects its latent factor, the greater the reliability value of the measurement. The magnitude of the construction reliability value can be calculated using the formula 1.

\[
CR = \frac{(\Sigma \text{SLF})^2}{(\Sigma \text{SLF})^2 + (\Sigma e)}
\]  

(1)

The second reliability test is done by calculating the average variance extracted (AVE). Explains that AVE shows the amount of variance from the indicators extracted by the developed latent construct. The magnitude of the construction reliability value can be calculated using the following formula 2.

\[
AVE = \frac{\Sigma \text{SLF}^2}{\Sigma \text{SLF}^2 + (\Sigma e)}
\]  

(2)

Hair, et al., (2010) stated the minimum recommended CR value is above 0.60 and the minimum acceptable AVE value is at least 0.50.

RESULTS AND DISCUSSIONS

Jeanrie and Bertrand (1999) explained that the assessment from expert reviewers showed two types of equivalence, namely linguistic equivalence and conceptual equivalence. The results of the assessment of the rating scale comparability and similarity from expert reviewers are then calculated to obtain the mean score. Sperber (2004) explains that the equivalence of statement items can be achieved if the statement item has a mean score of < 4. Based on the calcu-
The results of the mean score, the comparability and similarity levels show no item has a mean score of more than 4. This condition means that there is little difference in language on the adaptation scale and there is no significant meaning problem between the original-scale version of the statement items of the original scale version and the adapted version, although there may be differences in the terms used. More results can be found in Table 1.

Calculating the score of the I-CVI assessment results by expert reviewers is done by summing the rating value on each item given by the reviewers, then divided by the number of expert reviewers. Furthermore, CVI for scale or S-CVI is determined by calculating the average I-CVI, namely the number of I-CVI scores divided by the number of items overall. Polit et al. (2007) explained that an item is considered good if it has an I-CVI of at least 0.78 (based on the results of assessments with three or more expert reviewers) while if the item has an I-CVI below 0.78 then it should be removed or not used. A similar opinion was expressed by Zamanzadeh et al. (2015) that the removed item is an item that has an I-CVI below 0.70, while an item with I-CVI ranges from 0.70 - 0.78 can still be used with some revision.

### Table 1. Equivalence Results based on Expert Review

<table>
<thead>
<tr>
<th>No</th>
<th>Original Version</th>
<th>Back Translation Version</th>
<th>Mean Score</th>
<th>Comparability</th>
<th>Similarity</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I am able to negotiate and accomplish what is expected of me at work and in my family</td>
<td>I am able to negotiate and fulfill what is expected of me both at work and in my family</td>
<td>1.33</td>
<td>1</td>
<td></td>
<td>Equivalent</td>
</tr>
<tr>
<td>2</td>
<td>I do a good job of meeting the role expectations of critical people in my work and family life</td>
<td>I work well to fulfill the role that is expected by important people within my work and family life</td>
<td>1.66</td>
<td>1.66</td>
<td></td>
<td>Equivalent</td>
</tr>
<tr>
<td>3</td>
<td>People who are close to me would say that I do a good job of balancing work and family</td>
<td>People close to me will say that I can balance work and family well</td>
<td>1.33</td>
<td>1.33</td>
<td></td>
<td>Equivalent</td>
</tr>
<tr>
<td>4</td>
<td>I am able to accomplish the expectations that my supervisors and my family have for me</td>
<td>I can live up to the expectations that my boss and my family have for me.</td>
<td>1.66</td>
<td>2.33</td>
<td></td>
<td>Equivalent</td>
</tr>
<tr>
<td>5</td>
<td>My co-workers and members of my family would say that I am meeting their expectations.</td>
<td>My coworkers and family members will say that I lived up to their expectations.</td>
<td>1.33</td>
<td>1.33</td>
<td></td>
<td>Equivalent</td>
</tr>
</tbody>
</table>
| 6  | It is clear to me, based on feedback from co-workers and family members, that I am accomplishing both my work and family responsibilities | Based on the input provided by colleagues and family members, it is clear to me that I am fulfilling my responsibilities both at work and within the family. | 1.66       | 1.33          |            | Equivalent  

### Table 2. Recapitulation of calculations I-CVI

<table>
<thead>
<tr>
<th>N</th>
<th>Relevance</th>
<th>Importance</th>
<th>Clarity</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>verified</td>
</tr>
<tr>
<td>2</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>verified</td>
</tr>
<tr>
<td>3</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>verified</td>
</tr>
<tr>
<td>4</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>verified</td>
</tr>
<tr>
<td>5</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>verified</td>
</tr>
<tr>
<td>6</td>
<td>1,00</td>
<td>1,00</td>
<td>1,00</td>
<td>verified</td>
</tr>
</tbody>
</table>
sions or modifications. Based on the I-CVI assessment conducted in this study, a result of 1.00 was obtained which means that the I-CVI value is good. The overall results of the I-CVI calculation can be seen in Table 2.

Furthermore, S-CVI calculation is carried out by calculating the average I-CVI, which is the number of I-CVI scores divided by the total number of items. The S-CVI calculation resulted from 1.00. Polit et al (2007) stated that a good S-CVI is 0.90 and above. The S-CVI calculation in this study had a result of 1.00 so it can be concluded that the WFB scale has good content validity.

The first model testing of WFB scale adaptation with CFA first order obtained the results of the model that was already fit, but with modifications in accordance with the modification indices suggested Lisrel 9.10 obtained more fit results. The full results can be found in Table 3.

Based on the goodness of fit value above, it is known that the second model has a fit model criteria value that is more fit than the first model. Where the results of all criteria show indications of good fit i.e., RMSEA 0.062 below 0.08; SRMR 0.033 which is below 0.05; NFI and CFI values close to 1. The fit model means that the model/shape arranged is in accordance with existing data (empirical data). The second model is a modified result of the initial CFA model. Modification by connecting covariance from the error measurement values is recommended by Lisrel, namely error measurement between item number 3 and item number 2; and item number 6 with item number 1.

In addition, in confirmatory factor analysis (CFA), it is also known that all anchors of the WFB scale has an estimated factor loading in the range of 0.57 to 0.91. Hence it is concluded that the questions meet the valid criteria convergently in measuring the WFB construct. This is also reinforced by the Average Variance Extracted (AVE) value which states the convergent valid value on the construct (AVE 0.530) which is greater than 0.5. The entire WFB construct based on factor loading and AVE is concluded to meet the criteria of proportional validity. Evaluation of the validity of discriminant question items, is not carried out on a unidimensional scale, because it has only one dimension. Meanwhile, from reliability testing using Composite Reliability (CR) and Average Variance Extracted (AVE) values are known to be AVE values of 0.530 (> 0.5), and have a value of CR 0.87 (> 0.6).

Table 3. Comparison of the suitability of the initial and final CFA models of the WFB scale

<table>
<thead>
<tr>
<th>Goodness of Fit criterion</th>
<th>p value</th>
<th>RMSEA</th>
<th>SRMR</th>
<th>NFI</th>
<th>CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acceptable level of conformity</td>
<td>&gt; 0.05</td>
<td>≤ 0.08</td>
<td>≤ 0.05</td>
<td>≥ 0.9</td>
<td>≥ 0.9</td>
</tr>
<tr>
<td>Early model</td>
<td>0.00095</td>
<td>0.14</td>
<td>0.056</td>
<td>0.94</td>
<td>0.96</td>
</tr>
<tr>
<td>Final model</td>
<td>0.199</td>
<td>0.062</td>
<td>0.033</td>
<td>0.98</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Figure 3. Path Diagram
This indicates that the level of reliability in the construction of family work balance is quite high. The results of the full WFB scale validity and reliability test can be seen in Table 4.

Table 4. WFB Scale validity and reliability test results

<table>
<thead>
<tr>
<th>Dimension/Item</th>
<th>Factor</th>
<th>AVE</th>
<th>CR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aitem 1</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aitem 2</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aitem 3</td>
<td>0.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aitem 4</td>
<td>0.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aitem 5</td>
<td>0.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aitem 6</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CONCLUSIONS

Based on the results of the validity and reliability test, it was concluded that the WFB construct was measured/explained validly and reliably by six question items, where the validity value has a loading factor value ranging from 0.57 - 0.91. The reliable value has an AVE value of 0.530 and Composite Reliability (CR) of 0.87. The results of this study show that the WFB Scale is a valid and reliable measuring tool to measure work-family balance in the Indonesian context. With a lot of items, it will be one of the measuring instruments that researchers can choose to avoid saturation of research subjects because they have to fill in large quantities of scales. But keep in mind that this measuring instrument was only tested on a sample of women who worked as policewoman. Research with samples with other professions is highly recommended if the scale will be expanded in use.

REFERENCES


Adaptation and Validation of the Work–Family Balance Scale


