Internal Validation of the Warwick-Edinburgh Mental Wellbeing Scale: Rasch Analysis in the Indonesian Context

Dimas Aryo Wicaksono
Universitas Airlangga
aryo.wicaksono@psikologi.unair.ac.id

Adiyo Roebianto
Universitas Mercubuana
adiyo.roebianto@mercubuana.ac.id

Bambang Sumintono
Universiti Malaya
bambang@um.edu.my

Abstract

This study's purpose is to analyze the psychometric properties of the Indonesian version of the Warwick-Edinburgh Mental WellBeing Scale (WEMWBS) using the Rasch model measurement analysis to test its reliability of measurements, item fit statistics, rating scale functions, and differential item functioning. A total of 855 Indonesian samples completed 14-items in the Indonesian version of WEMWBS. The Rasch analysis indicates that the data fit the model, the item-person reliability is above 0.8, and the item-person separation is above the minimum requirement. All items in the instrument have met the statistic criteria indices in terms of outfit mean square as all item-measure correlation is above 0.4. The Likert rating scale analyses found the scale functioning properly and free from disordered thresholds. Lastly, DIF by gender was detected in three items. Further testing and refinement need to be conducted to ensure the reliability of the Indonesian version of WEMWBS.

Keywords: Mental Well-Being, WEMWBS, Validation, Rasch model, DIF.

Introduction

People have become more interested in discussing positive mental health, which according to The World Health Organization (2004) is a foundation for individual and community well-being. Positive mental health allows individuals to function effectively, realize their ability, cope with the stressor, be more productive at work, and contribute more to the community. Mental health can be seen as human capital and physical health as mental illness correlates with a high social and economic burden to society (Keyes, 2013).

The measurement of well-beingin Indonesia mainly utilizes Psychological (Eggleston et al., 2001; Rachmayani & Ramdhani, 2014) and Subjective well-being (Pandyaswargo et al., 2015) that tend

to measure two distinctive bodies of well-being. Moreover, Maulana, Khawaja, and Obst (2019) developed the Indonesian Well-being scale, which measures dimensions, namely (1) Spirituality, (2) Positive Social Relation, (3) Basic Needs, and (4) Acceptance. The study reported that this scale's measurement correlated significantly with other universal well-being scales such as SWLS, WEMWBS, and K10.

Most mental health inventory scales focus on aspects of mental disorder (Tennant et al., 2007; Stewart-Brown & Janmohammed, 2008) and other well-being scales, excluding affective-emotion aspects (Watson, Clark &Tellegen, 1988), cognitive-evaluative facets (Diener *et al., 1985*) and psychological functioning aspects (Ryff & Keyes, 1995). The Warwick-Edinburgh Mental Well Being Scale (WEMWBS), developed by Tennant *et al.* (2007), focused mainly on general positive mental health. The WEMWBS has been developed beyond its original version (Tennant et al., 2007) used in various countries,and translated into 25 different languages. The developed versions of the scale have been validated scientifically. Some of the validated translated versions of the scale include the version in Spanish (López et al., 2013), Portuguese (Santos et al., 2015), Mandarin (Ng *et al.,* 2014; Dong *et al.,* 2016), Norwegian (Karpavic˘iute and Macijauskiene, 2016; Ringdal *et al.*, 2017), French (Trouselard*et al.,* 2016), and Dutch (Ikink, Lamers and Bolier, 2012).The widespread use of this scale indicates that the scale has been recognized internationally as a measurement tool of mental well-being.

This article aims to explain the adaptation and internal validation process of the WEMWBS in the Indonesian context. There is no validity evidence of WEMWBS scale for Indonesian samples, particularly using RASCH measurement model.The Rasch model measurement approach was employed for an internal validation process. This research is expected to describe the development of measurement well-being in the Indonesian context, suggest an alternative scale based on positive mental health indicators,and providea more comprehensive perspective in explaining mental well-being, thus it can be expected this study can be as reference for research on mental well-being especially in Indonesia.

Literature Review

The terms 'well-being' and 'positive mental health' are often used interchangeably in academic and policy literature (Tennant et al., 2007). However, Ryan and Deci (2001) argue that the idea is misleading. They further suggest that the current research trendon well-being focuses on how individuals function fully in their environment, rather than discussing two general perspectives of well-being: hedonic and eudemonic approaches. Tennant et al. (2007) provide a scale to capture broad conceptions of well-being, including cognitive-evaluative, affective-emotional and psychological functioning. Stewart-Brown (2015) states that The Warwick-Edinburgh Mental Well-being Scale (WEMWBS) was designed to capture all aspects of well-being, including feeling and functionality. He further suggests that this scale could become an important tool for well-being measurement as it is easy to complete, brief, and well-validated (Stewart-Brown, 2015).

This study describes the adaptation and validity test of the WEMWBS in the Indonesian context. Indonesia has the largest Muslim population globally; thus, the influence of Islamic teaching and values, the spirituality aspects, tend to be included when measuring Indonesian people's well-being (Maulana, Khawaja, & Obst, 2019). For instance, Rahayu (2016) found that Indonesia's religious people tend to be happier than non-religious people. Lun and Bond (2013) also suggest that spirituality correlates positively with subjective well-being.

The WEMWBS has been used invarious countries and cultural contexts that require different measurement approaches to measure psychometric properties. The WEMWBS tested using psychometrics testing methods is valid and reliable as itsreliability index, Cronbach's Alpha is 0.89. Furthermore, except with the PANAS-NA (-0.55), the scale correlates positively with various well-being scales such as WHO-5 (0.77), SDHS (0.76), PANAS-PA (0.73), SWLS (0.72), and Scale of Psychological well-being (0.73) (Stewart-Brown & Janmohammed, 2008). The WEMWBS in Northern Ireland version is reported to have Cronbach alpha's index of 0.93. The Principal Component Analysis result shown that the single underlying factor explained 54% of the variance (Lloyd & Devine, 2012).

The Spanish version of WEMWBS, which CFA analysis indicated as a unidimensional scale (Lopez *et al.*, 2013), is reported to have high reliability (Cronbach's Alpha of over 0.90) and total-item correlation from 0.44 to 0.76. Similarly, the Brazilian Portuguese version of WEMWBS also has high reliability with Cronbach's Alpha of 0.89. This version has a moderate positive correlation with other mental well-being scales such as WHO-5 (r=0.59), WBMMS (r=0.69) (Santos, 2015). Taggart *et al.* (2012) reportedthat the version of WEMWBS used to measure the well-being of Chinese and Pakistani people who stay in The UK shows a single factor loading of 0.60-0.82 and 0.51-0.83, respectively, and Cronbach alpha index of 0.92 and 0.91 for Chinese and Pakistani participants. Trouselard *et al.* (2016) reported that among the samples (workers, students, and patients), the factor analysis of French WEMWBS indicates 40.12% of variance from the workers and students. This version has Cronbach's alpha ranged from 0.84 – 0.90, with a confidence interval of 95%.

The Chinese version of WEMWBS was given to two different samples: undergraduate nurse trainees (Dong *et al.,* 2016) and Chinese-speaking patients. The latter's well-being was measured by the short version of the WEMWBS (Ng *et al.,* 2014). The Chinese version of the WEMWBS used to measure undergraduate nurse trainees' well-being indicates a highCronbach's alpha index of 0.94, and item-total correlation ranges from 0.66 to 0.83. The explanatory factor analysis is consistent with the predicted scale structure, and all factor loadings are equal (above 0.40). The Principal Component Analysis also reveals one factor explained 59.45% of total variances (Dong *et al.,* 2016). Furthermore, the short Chinese version of WEMWBS used to measure Chinese-speaking patients' well-being indicates the Cronbach's alpha of 0.89, consistent with the English version of WEMWBS. The Principal Component Analysis identified that a single factor explained 61.1% of the total variance (Ng *et al.,* 2014).

The WEMWBS analysis on Rasch model measurement indicates that the model fits the short version of WEMWBS (seven items) (Bartram, Sinclair & Baldwin, 2013; Stewart-Brown *et al.,* 2009), applied to The UK Veterinary Profession (Bartram, Sinclair & Baldwin, 2013) and the Scottish Health Education Population Survey (Stewart-Brown *et al.,* 2009). Seven of the 14

items fit residuals above the recommended range of $\pm$ 2.5 logit score. Six items were removed from the scale. Those items are no. 5 (+5.81),"I've had energy to spare," no. 8 (-10.59),"I've been feeling good about myself," no. 10 (-6.22), "I've been feeling confident," no. 12 (+7.94), "I've been feeling loved," no. 13 (+2.99), "I've been interested in new things," and no. 14, "I've been feeling cheerful" (-9.61) (Bartram, Sinclair & Baldwin, 2013). This result shows direct single factor measurement and aligns with Stewart-Brown *et al.* (2009); thus, the model fit the short version of the WEMWBS. Moreover, the correlation between short and original versions of the WEMWBS is 0.95 (Stewart-Brown *et al.,* 2009). Lastly, Differential Item Functioning (DIF) analysis indicates that this scale is free from DIF in gender and age criteria (Bartram, Sinclair & Baldwin, 2013; Stewart-Brown *et al.,* 2009).

Method

The Warwick-Edinburgh Mental Wellbeing Scale Instrument

This study used the Indonesian translated version of The Warwick-Edinburgh Mental Wellbeing Scale (WEMWBS) to measure Indonesian respondents' mental health level. The scale of WEMWBS was translated into the Indonesian language by using the translation-back-translation process (Brislin, 1970). After the permission to translate the model in the original language into Indonesian was granted, the authors translated the instruments into Indonesian; then confirmed their equivalency by asking two independent translators, having Indonesian as the first language, for the back translation. The translators, having English studies and psychology background, reviewed the translation separately. The original version authors then confirmed the translation's equivalency with some recommendations for revisions. Specifically, the recommendations were related to alternative expressions, and clarifications of some statements such as "having much energy" (item no. 5) into "I feel energetic," and "intimate relationship" (item 9) become "I feel close with others."

The WEMWBS instrument, consisting of 14 positively phrased items covering both eudaimonic and hedonic aspects of well-being, measures positive mental health aspects (Lopez *et al.*, 2012). The Indonesian version is similar to the original version as it usesa 5-point Likert scale, from 1

(never) to 5 (always), and contains three basic demographic questions (i.e., age, gender, and education).

Research participants

This research employed a convenience sampling technique to collect the data. The participants involved in the validation of the instrument were 900 Indonesian samples randomly vary from their background. There is no specific occupational background as a samples of the study, samples categorization based on gender 548 respondents (64,1 percents) are male and 307 respondents (35,9 percents) are female. Some details about demographic background of the participants can be seen on Table 1. Concerning ethical consideration, the respondents' consent to participate in this study was sought before they filled the questionnaire manually (paper and pencil). The questionnaire's front cover explains that participation in the survey is voluntary and anonymous.The respondents completing the questionnaire had been fully aware of these details.

The data were entered into Microsoft Excel and checked by the Rasch measurement model software named Winsteps version 3.73 for data validation and cleaning. The procedure found that 30 respondents provided outlier responses (all maximum ratings). Misfit responses from 15 respondents were also found and removed. The analysis determined the data from 855 respondents could proceed for further analysis. The demographic profile of the Indonesian data in this study show in Table 1.

Table 1

Demographic Data of Respondent (N = 855)

Demographics		Frequency	Percentage (%)
Gender	Male	548	64.1%
	Female	307	35.9%
Age	< 25 years old	123	14.4%
	26 - 38 years old	459	53.7%
	39 – 55 years old	259	30.3%
> 50 years old		12	1.4%
	missing	2	0.2%
Education level	High school	409	47.7%
	Diploma	30	3.5%
Bachelor degree		241	28.3%
Master degree		146	17.1%
	Doctorate	17	2.0%
	missing	12	1.4%

The Rasch Measurement Model

The data collected from the WEMWBS questionnaire are ordinal; the frequency of responses appeared as odd-probability. With calibration of item difficulties and person abilities, as well as logarithm function, The Rasch measurement model converted raw ordinal data (Likert type data) into logit (log odd unit) data that has equal-interval type quality. The item logit assesses the overall fitness of the instruments at variable and item levels (psychometric attributes. Moreover, the person logit assesses each person's result, including its fit with statistics information (Linacre, 2012; Boone, Townsend, & Staver, 2014; Sumintono & Widhiarso, 2014; Bond & Fox, 2015). In particular, the Rasch model rating scale is suitable for the WEMWBS questionnaire assessing human opinion/perception/attitude (Englehard, 2013; Bond & Fox, 2015). Rasch analysis can determine item difficulty level (item calibration), detect item fit, and identify item bias (called differential item functioning or DIF) with precise measurement. The

Rasch measurement model generates a person logit that indicates the well-being of an individual.

Result

Reliability and Unidimensionality

A two-facet rating scale model of item and person, Rasch rating scale model, was employed to analyze 14 items of Warwick-Edinburgh Mental Well Being Scale (WEMWBS) given to 855 respondents. The items were centered at zero, allowing 'float' calibration of the respondents' mental well-being level. Table 2 shows that the mean measure (logit) of the item is 0.00 logit, with a relatively low standard deviation of 0.31, suggesting a narrow dispersion of measures across the logit scale in item difficulty level. The person logit mean is +2.67 logit, indicating that most respondents have good mental well-being. However, the standard deviation is 1.56, showing that the well-being level is widely-dispersed among the participants.

Table 2
Summary Statistics of Person and Items

	Person	**Item**
N	855	14
Measures		
Mean	2.67	0.00
Standard Deviation, SD	1.56	0.31
Standard Error	0.05	0.09
Outfit Mean Square		
Mean	1.00	1.00
SD	0.56	0.18
Separation	2.78	5.14
Reliability	0.89	0.96
Alpha Cronbach	0.91	
Chi-square (χ^2)	20550*	
Raw variance	47.5%	

* p < 0.01

The average outfit mean-square statistics of items and persons are at their expected value of 1. Moreover, the chi-square value is significant, indicating uniformed data that fit the model (Boone et al., 2014; Engelhard, 2013). Separation index (equal or more than three) and reliability (more than 0.67) statistics of items and persons also show high reliability (Fisher, 2004). Lastly, the raw variance value is more than 40%, indicating that the instrument has a high uni-dimensionality measuring respondents' hearth quality (Boone et al., 2016).

Table 3.
Statistics of Rating Scale Analysis

Category	Data Category Count Used	Frequency %	Average measure	Rasch-Andrich Threshold Measure	SE
1= never	34	<1%	-1.27	None	-
2 = rarely	357	3%	-0.25	-3.60	0.19
3 = sometimes	2559	21%	1.37	-1.38	0.06
4 = often	5422	45%	2.63	+1.30	0.03
5 = always	3596	30%	3.99	+3.69	0.02

Rating scale analysis

This study employed Likert-type rating criteria of frequency from score 1 ("never") to 5 ("always"). The separation statistics of the rating criteria are shown in Table 3.All of the categories in the Indonesian version of WEMWBS were chosen by more than ten individuals. The average person measured by each category moved up monotonically with the rating scale. The ideal distance value between the rating scale is 1.40 to 5.0 logit (Van Zile-Tamsen, 2017).
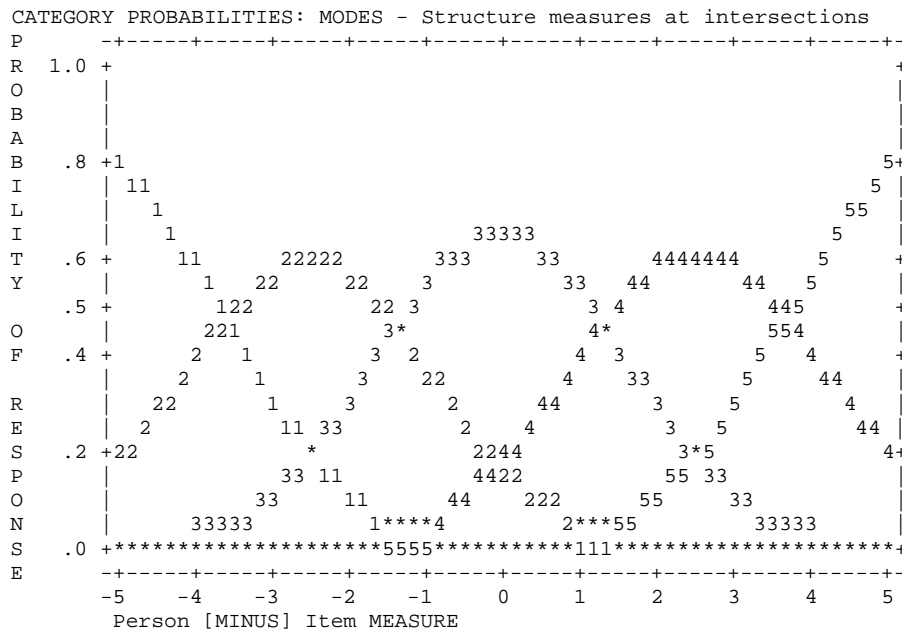
```
CATEGORY PROBABILITIES: MODES - Structure measures at intersections
P      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
R  1.0 +                                                               +
O      |                                                               |
B      |                                                               |
A      |                                                               |
B   .8 +1                                                            5+
I      | 11                                                        5 |
L      |    1                                                    55 |
I      |     1                          33333                  5   |
T   .6 +      11      22222       333      33       4444444      5     +
Y      |        1  22      22    3            33   44        44  5    |
    .5 +         122        22 3              3  4          445      +
O      |       221           3*               4*          554      |
F   .4 +      2    1         3   2           4   3        5    4      +
       |      2     1          3    22        4     33      5     44   |
R      |     22       1       3        2       44       3       5       4  |
E      |    2         11 33        2      4         3     5           44 |
S   .2 +22               *           2244             3*5               4+
P      |           33 11           4422           55 33             |
O      |          33        11      44      222        55      33       |
N      |      33333        1****4         2***55         33333     |
S   .0 +*******************5555**********111*******************+
E      -+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-
       -5    -4    -3    -2    -1     0     1     2     3     4     5
       Person [MINUS] Item MEASURE
```

Figure 1. Probability curves of response five Likert-type categories

Table 3 shows that the distances of the Rasch-Andrich threshold between the five rating categories were ideal. The result indicates that the respondents' rating scale was well-understood and effectively assessed respondents' mental well-being. Each unit rating scale's effectiveness was crucial for obtaining accurate and optimal results (Van Zile-Tamsen, 2017). Figure 1 shows that each rating probability curve has peaks and is separated, indicating that the rating is well-understood (Bond and Fox, 2015).

Table 4.
WEMWBS Item Calibration (N= 855)

Difficulty level	WWMBS Item	Logit Value Item	Stand. Error	Outfit MNSQ	Point Meas. Corr.
Rarely	W5	+0.50	0.06	1.06	0.62
	W3	+0.42	0.06	0.98	0.68
	W12	+0.36	0.06	1.17	0.62
	W6	+0.34	0.06	0.83	0.69
Sometimes	W7	+0.19	0.06	0.76	0.73
	W1	-0.03	0.06	1.40	0.62
	W2	-0.03	0.06	0.96	0.70
	W10	-0.07	0.06	0.81	0.75
	W9	-0.09	0.06	1.02	0.66
	W4	-0.14	0.06	1.21	0.64
	W14	-0.16	0.06	0.88	0.73
	W8	-0.30	0.06	0.87	0.73
Often	W11	-0.39	0.06	0.89	0.69
	W13	-0.61	0.06	1.15	0.62

Item Fit Statistics

Table 4 classifies the items according to their difficulty level or logit value (LVI). The items' classification into three difficulty levels was done by dividing the item logit score's distribution based on mean and standard deviation values (see Table 2). There were four items in the 'rarely" category (LVI > +0.31 logit), eight items in the 'sometimes' category (+0.31 $\geq$ LVI $\geq$ -0.31), and two items in the 'often' category (LVI < -0.31).

In terms of item fit statistics, outfit mean-square indices fall into fit criteria of 0.6 – 1.4. All measure correlation points have positive values and more than 0.4 showing all the statements are well-understood by the respondents (Engelhard, 2013; Boone et al., 2014). As shown by the standard error, the preciseness of measurement indicates proper indices (Bond and Fox, 2015).
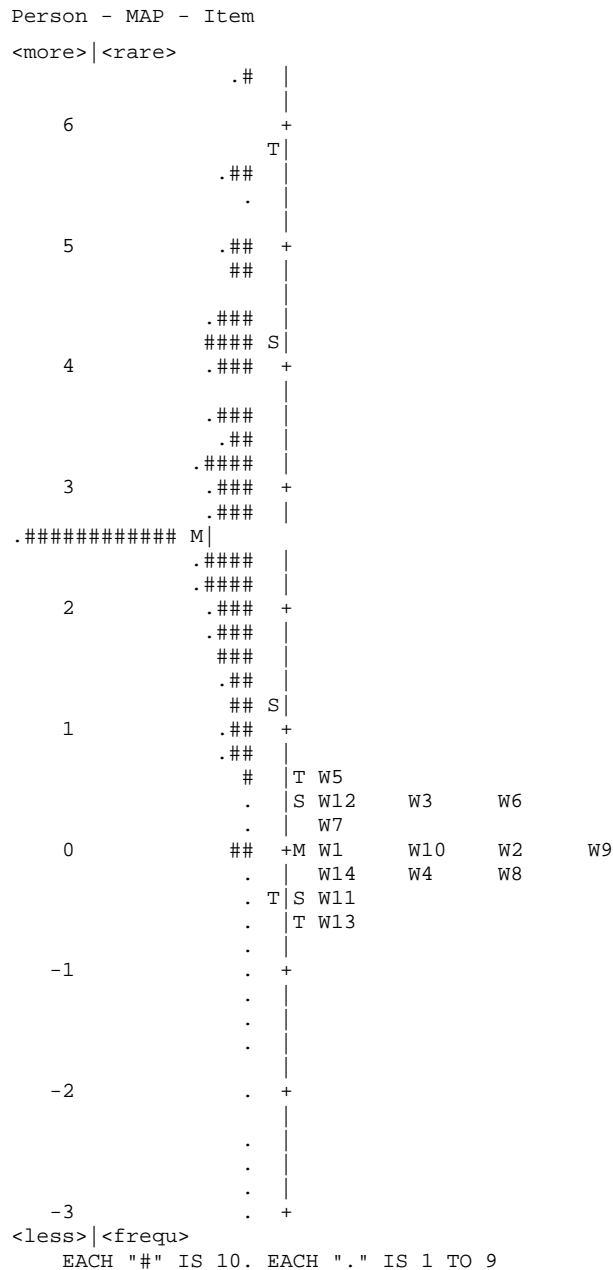
```
Person - MAP - Item
<more>|<rare>
                    .#  |
                        |
    6                   +
                      T |
                   .##  |
                    .   |
                        |
    5              .##   +
                    ##  |
                        |
                  .###  |
                  #### S|
    4             .###   +
                        |
                  .###  |
                   .##  |
                 .####  |
    3             .###   +
                  .###  |
.########### M|
                 .####  |
                 .####  |
    2             .###   +
                  .###  |
                   ###  |
                   .##  |
                   ## S|
    1              .##   +
                   .##  |
                    #  |T W5
                    .  |S W12     W3       W6
                    .  |  W7
    0              ##  +M W1      W10      W2       W9
                    .  |  W14     W4       W8
                  . T|S W11
                    .  |T W13
                    .  |
   -1               .   +
                    .  |
                    .  |
                    .  |
                        |
   -2               .   +
                        |
                    .  |
                    .  |
                    .  |
   -3               .   +
<less>|<frequ>
    EACH "#" IS 10. EACH "." IS 1 TO 9
```

Figure 2. Wright map of Indonesian version of the WEMWBS Calibration

Wright maps in Figure 2 display the joint depiction of respondents, items, and the well-being level. The person and item means (M) of 1 SD (S) and 2 SD (T) are respectively placed on the left and right sides of the line to facilitate map analysis. Figure 1 demonstrates that the difficulty of the items is lower than the participants' level of well-being. Moreover, the difficulty of the

items is between -0.61 to +0.5 logit, while the logit person mean of participants is +2.67 logit, indicating more respondents have a high-level of well-being.

DIF Analysis

Lastly, the different item functioning (DIF) analysis was conducted to examine the item bias of the demographic variables collected. The balanced group composition for DIF analysis, with more than 200 respondents for each group (see, for example, Boone et al., 2014; Bond & Fox, 2015), is shown in Table 2. The gender demographic data was then further analyzed.

Table 5.
Differential item functioning (DIF) analysis by gender in the WEMWBS items

Item	DIF		DIF	t	Prob.
	Male	Female	Contrast		
W8	-0.53	0.05	0.58	-4.75	0.0000
W10	-0.37	0.35	0.71	-5.90	0.0000
W12	0.72	-0.23	0.96	8.04	0.0000

Regarding gender, three items, W8, W10, and W12, are indicated to have DIF (Table 5). Zwick (2012) states that a slight to moderate DIF is present if the difficulty parameters in two different genders had significant probability and DIF contrast of more than 0.43. Figure 3 shows that male respondents tend to give more response than female respondents for item W8 (*I feel fine with myself*) and W10 (*I feel confident)*. On the other hand, female respondents tend to give more responses to item W12 (*I feel loved)* than male respondents.

Discussion

This study aims to validate the WEMWBS Indonesian version with the Rasch model measurement. Table 4 shows that the item fit statistics are in all good categories regarding to item difficulties, item fit statistics and outfit mean-square. All items'mean-square indices are in the acceptable fit range (0.6 – 1.4), and the point-measure correlationsare greater than 0.4. Table 3 and Figure 1 indicate that the Likert-type rating scale categories functioned well as all scales had satisfactory items and person reliability. The findings suggest a high probability that expected high-measure individuals or items would have greater scores (higher well being) than projected low-measure individuals or items. These psychometric properties report might have more accuracy than some previous research (see, for example, Bartram, Sinclair & Baldwin, 2013; Stewart-Brown et al., 2009). Unlike previous research (Bartram, Sinclair & Baldwin, 2013; Stewart-Brown et al., 2009) that took a specific group of individuals as their participants, this research selected the participants randomly and from various backgrounds. Random participant selection is important for the Rash Analysis to generate accurate results as it could be affected samplings heterogeneity (Cappelleri, Lundy, Hays, 2015).

Moreover, the Indonesian version of WEMWBS meets the unidimensionality assumption criteria since the identified constructs have been considered and effectively measured the latent trait. In the WEMWBS, the Rasch item residuals' PCA significantly proved the vulnerability of dependence on item-total correlation and Alpha Cronbach ofother measurement model construct validity.

The personand item separation in the translated version of WEMWBS is equal or more than 3, indicating the instruments' sensitivity to distinguish low and high trait level participants (Fisher, 2007). However, the scale items were most sensitive when measuring participants' traits at the moderate to low level. WEMWBS instrument development based on a Classical Test Theory (CTT) approach might account for this tendency, as the approach has two major issues that can affect validity: sets of redundant items and skewed response categories for most items (Petrillo,

Cano, McLeod, & Coon, 2015). This issue can be solved by raising item difficulty to increase separation index and adding more difficult items.

The mean for person measure in the scale was high, indicating that all items in the instrument tend to be too simplistic to be utilized in Indonesian samples. Similarly, research in cultural psychology also reported this finding. In the Brazilian Portuguese context, the report suggested that the WEMWBS was easy to use and comprehensive to measure mental well-being. The WEMWBS mean score in Brazilian contexts was 51.8±8 points, classified as a moderate level of mental well-being compared with the UK norm (Santos *et al.*, 2015). Moreover, research on Spanish samples found that the mean score of the WEMWBS was 53.5 (Lopez *et al.*, 2013), while research on Chinese samples found the mean score of 51.48±9.11 points (Dong *et al.,* 2016). In the English context, the WEMWBS norm results showed that socioeconomic factors relate to the mental well-being level. People with higher income and education tend to have higher mental well-being scores than people with lower income and education levels (Stewart-Brown *et al.,* 2015).

Similarly, in the Indonesian cultural context, well-being is quite related to financial well-being. The term 'sejahtera' or 'prosperous' is commonly referred as achieving financial well-being (Maulana, Khawaja, & Obst, 2019). Rahayu (2016) further suggests that an increase in total income and higher education can lead to more happiness for Indonesian people. Besides, Maulana, Khawaja, and Obst (2019) add that Indonesian people's acceptance or complacency plays an important role to the well-being. These factors can be the main explanations why Indonesian samples tend to have high well-being levelwhen measured by the WEMWBS.

DIF analysis showed that 11 items do not have statistical significance. However, this finding might not affect the measurement as The Rasch Measurement, focusing more on psychometric attributes, consider the significance test less influential. Three items indicated bias, possibly not because of the item quality, but by Indonesian respondents' typical response. For instance, female respondents tend to be more passionate with attention from others (item W12),

whereas male respondents, when looking to themselves, tend to be more assertive (item W8 and W10). This tendency may indicate that the three items in this scaleare not invariant measurements for males and females. The analysis suggests that the Indonesian version of WEMWBS scores between males and females are not generalized. Nevertheless, the gender variable should be taken into accounts to minimize gender bias when developing instruments.

Conclusion

The research concludes that the Indonesian version of WEMWBS scale had shown psychometric properties that significantly fit the Rasch model. The findings are different from the WEMWBS scale study that employed the Rasch model, especially in terms of participants' mental well-being score distribution. Table 2 shows that the mean of mental well-being logit scores was 2.67, above a zero (0) scale. The findings indicate that the participants in this research have a relatively high mean score of mental well-being.

Even though the scale fit the Rasch model, three items, W8 ("*I feel fine with myself*"), W10 ("*I feel confident*"), and W12 ("*I feel being loved*"), applied differently for male and female participants. The first two items, W8 and W10, seem to be favored more by male than female participants. Meanwhile, female participants tend to be more open to W12 item. Further research employing a qualitative approach, such as interviews, needs to be conducted to reveal the influence of gender on these three items. DIF analysis can also be conducted based on other potentially influential variables such as age, ethnicity, and education level. Moreover, a qualitative approach needs to be employed to reveal to what extent participants' backgrounds relate to the well-being scores. Further research on mental well-being should employ a mixed method to reveal a deeper and more compelling finding.

References

Bartram, D. J., Sinclair, J. M., Baldwin, & David S. (2012). Further validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in the UK veterinary profession: Rasch Analysis. *Qual Life Res,* 22: 379-391.

Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences.* Third Edit. New York: Routledge.

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences.* New York: Springer.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology*, 1(3) September 1970, pp. 185-216.

Cappelleri, J. C., Jason Lundy, J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, 36(5), 648–662. https://doi.org/10.1016/j.clinthera.2014.04.006.

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, 49:1, 71-75, DOI: 10.1207/s15327752jpa4901_13.

Dong, H. (2016). Translation and validation of Chinese version of the Warwick-Edinburgh Mental Well-being Scale with Undergraduate nursing trainees. *Journal of Psychiatric and Mental Health Nursing*. 23, 554-560.

Eggleston, E. (2001). Measuring women's psychological well-being in Indonesia. *Women & Health,* 32(4), 156-162.

Engelhard, Jr. G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences.* New York: Routledge.

Fisher, W.P. Jr. (2007). *Rating scale instrument quality criteria. Rasch measurement transactions,* 21(1) p. 1095 (available at http://www.rasch.org/rmt/rmt211m.htm).

Ikink, J. G. M., Lamers, S. M., & Bolier, J. M. (2012). De Warwick-Edinburgh Mental Well-being Scale (WEMWBS) alsmeet instrument voor mentaal welbevinden in Nederland. *Master thesis.*

Karpavicˇiūtė, S & Macijauskienė, J. (2016). The impact of arts activity on nursing staff well-being: an intervention in the workplace. *International Journal of Environmental Research and Public Health*. 13. 435.

Keyes C.L.M. (2013) *Promoting and protecting positive mental health: early and often throughout the lifespan*. In: Keyes C. (eds) Mental Well-Being. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-5195-8_1

Linacre, J.M. (2012). *Many-facet Rasch measurement: Facet tutorial*. Retrieved from http://www.winsteps.com/a./ftutorial2.pdf

Lloyd, K., & Devine, P. (2012). Psychometric properties of The Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in Northern Ireland. *Journal of Mental Health.* June 2012, 21(3), 257-263.

Lopez, M. A. (2013). Adaptation into Spanish of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) and preliminary validation in a student sample. *Qual Life Res,* 22: 1099-1104.

Mangkualam, K. (2018) The relationship between independent behavior and positive mental health toward entrepreneurial intention of psychology students. *International Postgraduate Research Conference*, October 2018.

Maulana, H., Khawaja, N., & Obst, P. (2019). Development and validation of The Indonesian Well-Being Scale. *Asian Journal of Social Psychology*, 22, 268-280.

Ng, S.S.W. (2014). Translation and validation of the Chinese version of the Short Warwick-Edinburgh Mental Well-being Scale for patients with mental illness in Hongkong. *East Asian Arch Psychiatry,* 24:3-9.

Pandyaswargo, A. H., Abe, N., & Fransiscus, Y. (2015) Objective and subjective measurement for development and well-being: some perspectives from a bottom-up study in Indonesia. *Global Environmental Research*. 19/2015: 207-216.

Petrillo, J. (2015) Patient-reported outcomes using classical test theory, item response theory and rasch measurement theory to evaluate patient-reported outcome measures: a comparison of worked examples. *Value in Health*, 18: 25 – 34.

Rahayu, T. P. (2016) The determinants of happiness in Indonesia. *Mediterranean Journal of Social Science,* 7(2), 234-239.

Rachmayani, D., & Ramadhani, N. (2014) Adaptasi bahasa dan budaya skala psychological well being. *Proceeding Seminar Nasional Psikometri.* Universitas Muhammadiyah Surakarta.

Ringdal. R. (2017). Validation of two versions of the Warwick-Edinburgh Mental Well-being Scale among Norwegian adolescents. *Scandinavian Journal of Public Health.* 1-8.

Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: a Review of research on hedonic and eudaimonic well-being. *Annual Rev Psych*. 52:141-66.

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69*(4), 719–727. https://doi.org/10.1037/0022-3514.69.4.719.

Santos, J. (2015). Adaptation and cross-cultural validation of the Brazilian version of the Warwick-Edinburgh mental well-being scale. *Rev Assoc Med Bras,* 61(3), 209-214.

Seasia, (2017), *Indonesian young people have the best mental well-being in the world*. https://seasia.co/2017/02/11/indonesian-young-people-have-the-best-mental-wellbeing-in-the-world.

Sumintono, B. & Widhiarso, W. (2014). *Aplikasi model Rasch untuk penelitian ilmu-ilmu sosial (edisirevisi) [Application of Rasch modelling in social science research, revised edition]*. Cimahi: Trimkom Publishing House.

Stewart-Brown, S. (2015). Measuring Well-Being: What does the Warwick-Edinburgh Mental Well-being Scale have to offer integrated care? *European Journal of Integrative Medicine,* Volume 7(4), 384-388.

Stewart-Brown, S. (2009). Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. *Health and Quality Life Outcomes,* 7:15.

Stewart-Brown, S., & Janmohamed, K. (2008), Warwick-Edinburgh Mental Well-being Scale (WEMWBS) User Guide. *Warwick Medical School, University of Warwick*.

Taggart, F. (2013). Cross-cultural evaluation of the Warwick-Edinburgh mental well-being scale (WEMWBS)- a mixed methods study. *Health and Quality of Life Outcomes,* 11:27.

Tenant, R. (2007). The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): development and UK Validation. *Health and Quality of Life Outcomes*, 2007, 5:63.

Trousselard, M. (2016). Validation of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in French psychiatric and general populations. *Psychiatric Research.* 245 (2016) 282-290.

Van Zile-Tamsen, C. (2017). Using Rasch analysis to inform rating scale development. *Research in Higher Education,* 58(8), 922-933.

Watson, D., Clark, L. A., & Tellegen, A. (1988) Development and Validation of Brief Measure of positive and Negative Affect: The PANAS Scale, *Journal of Personality and Social Psychology*. Vol.54, No. 6, 1063-1070.

WEMWBS, **"***WEWMWBS In Different Language,"*https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/using/translations/

WHO. (2004) *Promoting mental health: concepts, emerging evidence, practice*: *Summary report.* Geneva, World Health Organization.

Zwick, R. (2012) A Review of ETS Differential item functioning assessment procedures: flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report Series.*