Asia-Pacific Social Science Review 19(3) 2019, pp. 247–254

RESEARCH BRIEF

Rasch Analysis and Differential Item Functioning of Work-Related Flow Inventory (WOLF): A Study on the Hotel Industry in Sarawak

Mark Kasa, * Donald Stephen, and Zaiton Hassan
Universiti Malaysia Sarawak
*markyek@yahoo.com

This study has examined the psychometric properties of Work-Related Flow Inventory (WOLF) that examines the flow experience at work by measuring employees' absorption, enjoyment, and intrinsic motivation (Bakker, 2005, 2008). Absorption is defined as a state of total concentration while immersed in the work. As for enjoyment, it is referred to as a positive judgment on the matter of the quality of work. Lastly, intrinsic motivation is defined as the willingness to exhibit certain behaviors that stemmed from inherent satisfactions rather than separable consequences (Ryan & Deci, 2000). The term *flow* was introduced by Csikszentmihalyi (1990) who defined it as "the state in which people are very intensely immersed in an activity that nothing else seems to matter; the experience itself is so enjoyable whereby people will do it even at a great cost, for the sheer sake of doing it" (Csikszentmihalyi, 2003, p. 4). It was recognized that the notion of flow has a positive relationship with the productivity of an organization. The notion of flow is attributed to the willingness to seek constant optimal experiences that could lead to more willingness to invest time and effort in a task voluntarily (Seligman, Csikszentmihalyi, Fredrikson, Wateman, & Emmons, 2011).

This study focuses on the hotel industry due to its competitiveness and stressful nature attributed to frequent strenuous labors such as long and irregular working hours, midnight shift, and working nature during holiday seasons (Kasa & Hassan, 2015). To overcome issues pertaining to work stressors, studies revealed that an organization could cultivate and boost productivity through the experience of flow (Csikszentmihalyi, 2003). The statement supported the point that the working environment in which the flow can be experienced by the employees has the possibility to encourage their development and heighten productivity, as well as increasing organizational productivity.

There is no published study found on WOLF validation that utilizes homogenous samples. The use of homogenous samples can eliminate unwanted variability due to the influence of sociodemographic factors. Noises associated with nuisance sociodemographic factors can hinder correct judgment about the population of interest (Bornstein, Jager, & Putnick, 2013). According to Bakker (2008), WOLF validation studies derived its data from various occupational groups and companies in the Netherlands, which includes TV stations, temporary work agencies, self-employed trainers, insurance companies, businesses, as well as health providers. There is no specific WOLF validation study conducted in the hotel industry. There is a debate on the use of WOLF in the hotel industry as the existing instrument was the result of validation studies derived from data sources which are not

relevant to the context of use. Due to this concern, it is crucial to re-examine the psychometric properties of WOLF and assess the suitability of the instrument to be used in the hotel industry.

The WOLF validation studies in the Asian industries are also scarce. Among the notabilia of WOLF studies in Asia are by Chen, Ye, Chen, and Tung (2010) in Taiwan; Moneta (2004) in China; Soltani, Roslan, Abdullah, and Jan (2011) in West Malaysia; Paulus (2012) in Indonesia; and Kasa and Hassan (2015, 2016, 2017) in Sarawak, Malaysia. The implementation of WOLF in the studies conducted in Asia was developed and validated based on the Western context. The aspect of localization and cultural difference have been regarded as nuisance variables by researchers although it has been documented that cultural difference and sociodemographic characteristics can affect a study's outcomes (Bornstein et al., 2013).

Past studies in Europe and Africa have mentioned that WOLF demonstrated high reliability, factorial, constructive, as well as predictive validity (Bakker, 2008; Bakker, Golub, & Majdarijavec, 2017; Geyser, Geldenhuys, & Crous, 2015). However, the past studies on WOLF validation were conducted using classical measurement methods, namely, exploratory and confirmatory factor analysis (EFA & CFA). This limitation is the reason that this study has chosen to validate the instrument using Andrich's rating scale model, which is an extension of the Rasch model. The Rasch model is considered as a modern measurement method. Unlike CFA, parameters derived from the Rasch measurement model are independent of the sample being used (Ewing, Salzberger, & Sinkovics, 2009). This characteristic can be put to good use when dealing with homogenous samples.

Methods

This study uses convenience sampling, which made up of 250 full-time hotel employees—141 females and 109 males. The respondents' ethnic backgrounds are as follows: Chinese (50.8%), Malay (19.2%), Iban (9.6 %), Bidayuh (8.0%), and others (8.8%). The age groups of the respondents are as follows: 21–30 years old (56.4%) and 31–40 years old (22.8%). A total of four hotels in Kuching, Sarawak (three from 3-star hotels and one from 4-star hotel) have participated in this study with prior approval given to the researcher. The hotel employees have filled out the questionnaires voluntarily and anonymously, which were managed by their respective human resources manager. The process took approximately 12 weeks to complete the data collection.

WOLF includes 13 items that measure the construct of absorption (four items), work enjoyment (four items), and intrinsic work motivation (five items). The respondents were asked to indicate the frequency of their experience for the previous week (1 = never, 6 = everyday).

There is an abundance of psychometric techniques that can be used to develop a measure for a study. Rasch analysis was developed to improve the quality and precision of a measure (Boone, 2016). This study adopts the Andrich's (1978) rating scale model because all items share the same response anchors that also share a common threshold across all items. The use of Andrich's rating scale model is suitable for polytomous items as it will improve the estimation of thresholds (Meyer, 2014). For Likert-type item j with h=1, …, response categories, given that θ, b_j, and τ_v represent person ability, item difficulty, and threshold parameter respectively, the probability of the response k is given by the following formula:

$$P_{jk}(\theta) = \frac{exp\sum_{v=1}^{k}\left(\theta - b_j - \tau_v\right)}{\sum_{h=1}^{m_j} exp\sum_{v=1}^{k}\left(\theta - b_j - \tau_v\right)}$$

The item invariance in Rasch analysis refers to the independence of estimated item location parameters to the sample in which the estimates are derived from. It is believed that an item does not demonstrate item invariance if it exhibits item bias or differential item functioning (DIF) (Smith, Wright, Selby, & Velikova, 2007)in particular using non-sample dependent measurement techniques, such as Rasch Models. Furthermore, few studies have explored the relationship between item fit to the Rasch Model and clinical utility. The aim of this study was to investigate the dimensionality and measurement properties of the FACT-G with Rasch Models and Factor analysis.\\n\\ nMETHODS: A factor analysis and Rasch analysis (Partial Credit Model. The Mantel–Haenszel's chi-square test was used to compare the tendency of selecting scale response anchors from the two groups of individuals that match on a person's ability logits on the continuum of the latent trait. The item's characteristic curves with Nadaraya–Watson kernel regression

estimator was adapted to visualize the presence of differential item functioning in any part along the continuum of a latent trait. The curves were smoothed by a Gaussian kernel with a bandwidth value of 1. The assessment of DIF classified the items into three categories: AA (negligible DIF), BB (medium DIF), or CC (large DIF). If an item favors the focal group, a plus (+) sign is added to the classification, and if an item favors the reference group, a minus (-) sign is added (Cunningham et al., 2011; Meyer, 2014). This study has used R Statistical Software version 3.4.2, JASP, and JMetrik version 4.1 to create all graphs and to perform the statistical analyses.

Results

Reliability

Referring to the item and person reliability in Table 1, it can be concluded that the items in all subscales are reliable (item reliability = 0.7657-0.9850; person reliability = 0.8146- 0.8801). The range for item reliability confirms that the WOLF in the hotel industry exhibits a good level of reliability. Separation index is used to explain the number of distinguishable groups from the measurement (Wilson, 2005).

To be considered as reliable, the item separation index should be more than 3 and the person separation index should be no less than 2 (Linacre, 2012). The WOLF items collectively show good level of separation index (item separation = 6.3074, person separation = 3.0963). The values prove that the person or items that were being estimated to have a higher score achieve a higher score. It was found that only "enjoyment" shows a slightly low value for item separation (1.8079).

Rasch Analysis

Table 2 shows the values for item discriminant (Disc.), overall Rasch difficulty (Diff. Overall), Rasch difficulty based on subscales using joint maximum likelihood estimation (Diff. JMLE), weighted mean squares (WMS), unweighted mean squares (UMS), standardized weighted mean squares (Std. WMS), and standardized unweighted mean squares (Std. UMS).

There are various cut-off points for fit statistics due to the variability of contexts and risks. The advised range for mean squares value for productive measurement for low-stake rating scale is from 0.50 to 1.50, and the standardized mean squares should be in the range of -2 to 3 (Linacre, 2002). A measure is considered as unidimensional if all the items have acceptable fit statistics (Baghaei, 2008). Besides that, the items of all subscale are recorded as good weighted and unweighted mean squares (WMS= 0.57-1.30, UMS= 0.58-1.30). Other than Item 11 (Std. WMS= -5.43, Std. UMS= -5.35) and Item 3 (Std. WMS= -2.57, Std. UMS= -2.63), the remaining items exhibit non-extreme values of standardized weighted and unweighted mean squares (Std. WMS= -1.93-2.99, Std. UMS= -2.10-2.97). Although item 11 and item 3

Table 1

Rasch Reliability for WOLF and its Subscales

Statistic	Overall		Absorption		Enjoyment		Motivation	
	Items	Persons	Items	Persons	Items	Persons	Items	Persons
Observed Variance	0.2135	1.1981	0.1279	2.8115	0.0589	6.5828	0.3573	1.5996
Observed Std. Dev.	0.4620	1.0946	0.3577	1.6767	0.2427	2.5657	0.5977	1.2648
Mean Square Error	0.0052	0.1132	0.0081	0.5214	0.0138	0.7891	0.0054	0.2922
Root MSE	0.0724	0.3364	0.0898	0.7221	0.1175	0.8883	0.0733	0.5406
Adjusted Variance	0.2082	1.0849	0.1199	2.2901	0.0451	5.7938	0.3519	1.3074
Adjusted Std. Dev.	0.4563	1.0416	0.3462	1.5133	0.2123	2.4070	0.5932	1.1434
Separation Index	6.3074	3.0963	3.8563	2.0959	1.8079	2.7097	8.0953	2.1151
Number of Strata	8.7432	4.4618	5.4751	3.1278	2.7439	3.9463	11.1270	3.1535
Rasch Reliability	0.9755	0.9055	0.9370	0.8146	0.7657	0.8801	0.9850	0.8173

Table 2

Item Discriminant and Fit Statistics

Components	Item	Disc.	Del.	Diff. Overall	Diff. JMLE	WMS	UMS	Std. WMS	Std. UMS
Absorption	f01	0.42	0.89	0.29	0.21	1.07	1.09	0.81	1.00
	f02	0.41	0.89	0.29	0.21	0.84	0.84	-1.77	-1.81
	f03	0.58	0.89	0.28	0.19	0.78	0.77	-2.57	-2.63
Enjoyment	f04	0.59	0.88	-0.28	-0.62	1.30	1.30	2.99	2.97
	f05	0.72	0.88	-0.32	0.23	1.18	1.17	1.79	1.67
	f06	0.73	0.88	-0.41	0.10	0.92	0.90	-0.78	-1.05
	f07	0.77	0.87	-0.48	-0.08	0.82	0.81	-1.93	-2.10
	f08	0.70	0.88	-0.59	-0.34	1.05	1.03	0.53	-0.31
Motivation	f09	0.41	0.89	1.02	0.91	1.25	1.24	2.77	2.52
	f10	0.50	0.89	0.66	0.49	1.04	1.11	0.51	1.26
	f11	0.73	0.88	-0.02	-0.32	0.57	0.58	-5.43	-5.35
	f12	0.61	0.88	-0.13	-0.44	0.93	0.92	-0.72	-0.80
	f13	0.67	0.88	-0.30	-0.64	1.12	1.09	1.30	0.91

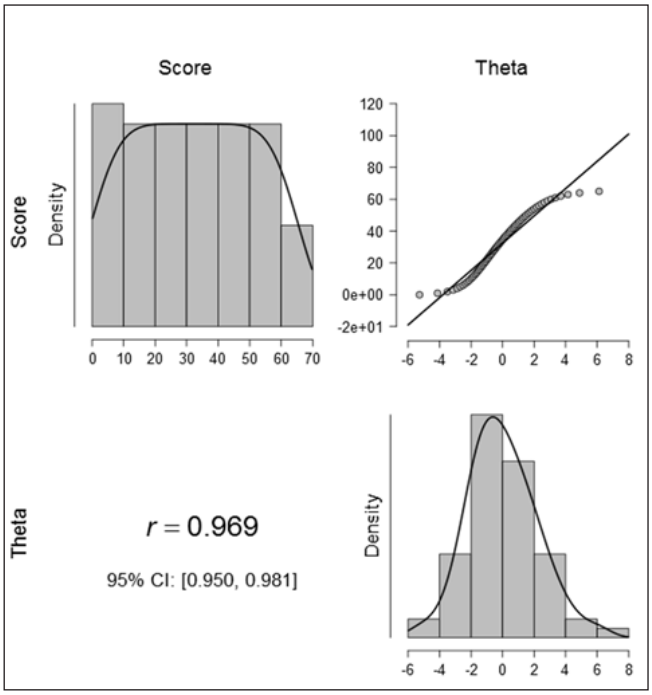

Figure 1. Density and test characteristic curve.

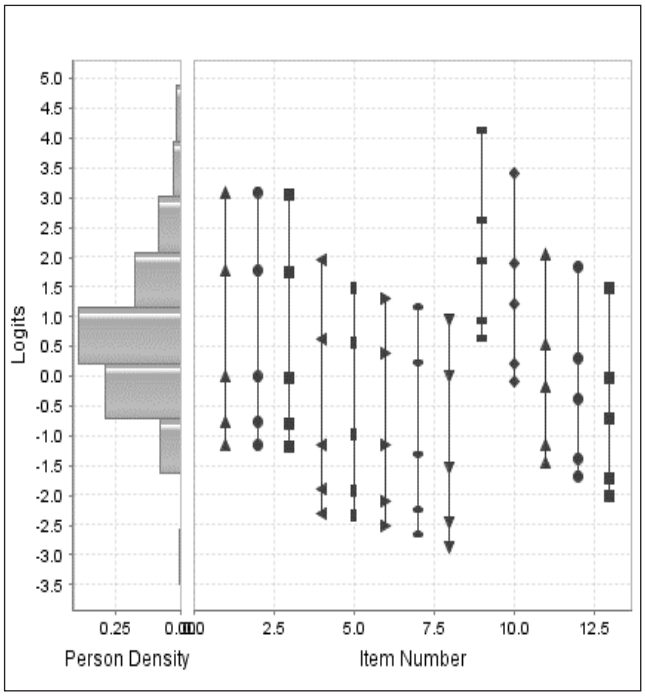

Figure 2. Item map.

exhibit high predictability, there is not enough evidence to remove these items as their mean square values are deemed productive for measurement, and the items do not degrade the measure.

Figure 1 shows the raw scores density and person ability density. Rasch produces the estimation person ability that corresponds to each score. Based on this information, the theoretical relationship between person ability estimates (theta) and raw scores was estimated (r=0.969 at 95% Confident Interval [0.950, 0.981]) to produce the test characteristic curve. The theta density in Figure 1 is analogous to the theta density of the item map in Figure 2. Figure 2 also presents the ability of each item to cover a sufficient range of person ability. Referring to Figure 2, item 9 is the most difficult item (higher level of flow needed for endorsement), whereas item 8 is the easiest item (lower level of flow needed for endorsement). Overall, the item alignment is good because most items are in the high-density region of person distribution. The item map also shows that there is a more extreme person compared to extreme items (theta= -5.2751 to 6.1251).

Differential Item Functioning

This study has appointed the male participants as the focal group and the female participants as the reference group for the analysis of DIF. It was conducted to examine whether the estimated item location parameters are dependent on the participant's gender. Referring to Table 3, item 1 (=6.45, p=0.01, effect size/common odds ratio=0.33) and item 9 (=3.06, p=0.08, effect size=-0.01) show moderate DIF. The common odds ratio values of these items are small, and they could not be deleted. However, both items should be further investigated in future studies as they are non-uniform DIF. Figure 3 shows the probability density of responses for the male and female groups. Referring to Figure 3, it is shown that item 1 and item 9 are non-uniform DIF. The non-uniform pattern is the reason that these two items need to be revised to minimize the bias associated with gender.

Table 3

DIF Analysis of WOLF

Item		p	E.S. (95% C.I.)	Class
(1) I think about nothing else when I am working.	6.45	0.01	0.33 (0.05,0.61)	BB+
(2) I get carried away by my work.	0.04	0.85	0.02 (-0.25,0.28)	AA
(3) I forget about everything else around me when I am working.	0.15	0.70	-0.05 (-0.29,0.18)	AA
(4) I am totally immersed in my work.	1.11	0.29	-0.09 (-0.31,0.12)	AA
(5) My work gives me a good feeling.	3.75	0.05	-0.33 (-0.33,0.06)	AA
(6) I do my work with much enjoyment.	0.19	0.67	-0.22 (-0.22,0.15)	AA
(7) I feel happy during my work.	0.07	0.79	-0.19 (-0.19,0.17)	AA
(8) When I am working, I feel cheerful.	1.74	0.19	-0.35 (-0.35,0.05)	AA
(9) I would still do this work even if I receive less pay.	3.06	0.08	-0.1 (-0.01,0.67)	BB+
(10) I find out that I also want to work during my free time.	0.76	0.38	-0.16 (-0.16,0.43)	AA
(11) work because I enjoy doing it.	1.55	0.21	-0.28 (-0.28,0.09)	AA
(12) I am doing what I do for myself when I am working on something.	1.70	0.19	-0.41 (-0.41,0.07)	AA
(13) gain my motivation from the work itself and not for the reward.	0.05	0.82	-0.26 (-0.26,0.17)	AA

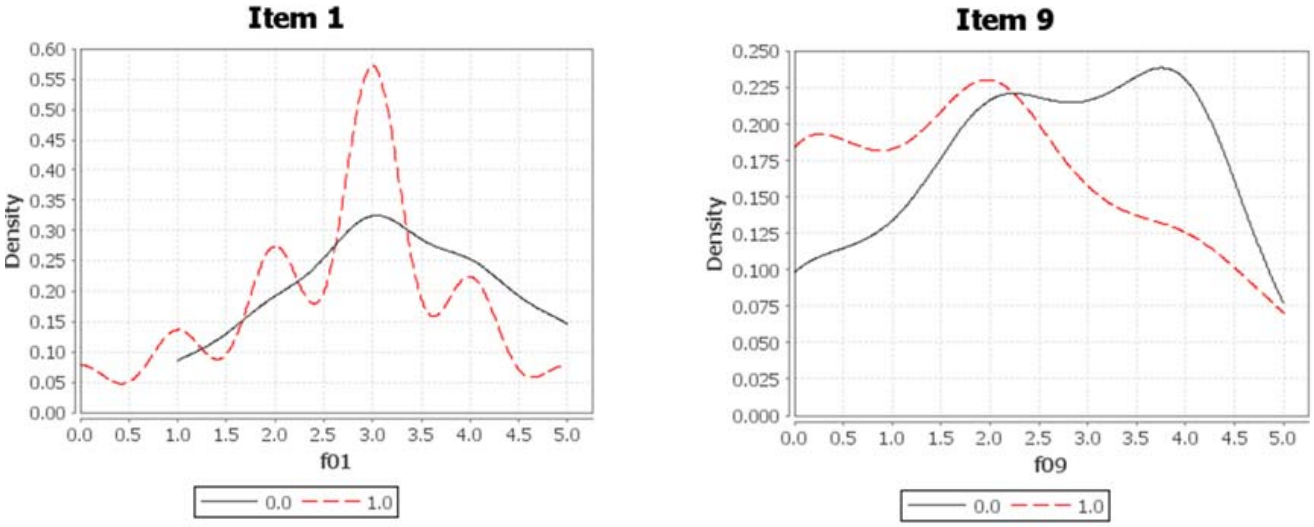

Figure 3. Probability of responses for item 1 and item 9 in terms of gender.

Discussion

This study uses the Rasch reliability to examine the inter-scale consistency as its indicators provide more information about the measure compared to classical reliable measures such as Cronbach's alpha (Aziz, Masodi, & Zaharim, 2013). In this study, the overall person reliability of 0.9055 is consistent with the result of classical reliability coefficients (Cronbach's alpha = 0.8957, Guttman's Lambda 2 = 0.9004, Feldt-Gilmer = 0.8970, Feldt-Brennan = 0.8970, Raju's Beta = 0.8957). As a comparison, there is a study conducted by the developer of WOLF that reported the Cronbach's alpha values for all the subscales are in the range from 0.81 to 0.87 (Bakker et al., 2017). The cross-validation between Rasch, various classical reliability coefficients, as well as findings from previous research have confirmed the high reliability of the WOLF instrument.

For the factor structure, past researchers have adopted CFA to examine the unidimensionality of WOLF instrument. A study reaffirmed that the original stance of a three-factor model (absorption, enjoyment, and intrinsic motivation) is the best fit for WOLF instrument but also mentioned the existence of two-factor model (absorption and work enjoyment/ motivation; Geyser et al., 2015). The WOLF developer subsequently extended the use of WOLF into a learning environment with the conception of the Study-Related Flow Inventory (WOLF-S) and reiterated the three-factor model (Bakker et al., 2017). Despite the conceptual differences between classical and modern test theory, it was noted that the results could be similar in some cases. Another difference is the nature of both approaches that tackle the issue of reliability differently (Demars, 2013; Wolfe & Singh, 2011)results are compared from a confirmatory factor analysis (CFA. The Rasch model also provides more information about the measure by assessing invariance, illustrating the relationship between person ability and item difficulty with the construct map, as well as interpreting and targeting measures with reference to items (Christensen, Engelhard, & Salzberger, 2012). Overall, this study has explained about the measure that has not been assessed in past research.

This study has several limitations regarding the prevalence of WOLF validation. The respondents were not randomly selected from the entire full-time hotel employees' population, and the method has some selection bias criteria that could affect the results. However, future studies may be conducted to determine the robustness of the findings in other samples to retrieve better samples for the other regions in Sarawak.

This study has tackled the problem statements and gaps by investigating the psychometric properties of WOLF in the hotel industry. In conclusion, the validation that uses Rasch analysis has confirmed that WOLF can be applied in the hotel industry and reaffirmed the unidimensionality of each subscale. Besides that, WOLF in the context of the hotel industry

has also exhibited high reliability. Mantel–Haenszel test has detected two items (item 1 and item 9) that show a slight dependency of estimated item location parameters on the participant's gender. The outcome of this study can be put into good use for future implementation of WOLF.

Acknowledgment

This project is funded under Post-Doctoral Grant F04/DPD/1626/2017 by Universiti Malaysia Sarawak.

Declaration of ownership

This report is our original work.

Conflict of Interest

None

Ethical clearance

The study was approved by the institution.

References

Andrich, D. (1978). Rating formulation for ordered response categories. *Psychometrika, 43*(4), 561–573.

Aziz, A. A., Masodi, M. S., & Zaharim, A. (2013). Asas model mengukuran Rasch: *Pembentukan skala dan struktur pengukuran* [Introduction to Rasch measurement model: Scale development and Measurement Structure]. Bangi: Penerbit Universiti kebangsaaan Malaysia.

Baghaei, P. (2008). The Rasch model as a construct validation tool. *Rasch Measurement Transactions, 22*(1), 1145–1146. Retrieved from http://www.rasch.org/rmt/rmt221a.htm

Bakker, A. B. (2005). Flow among music teachers and their students: The crossover of peak experiences. *Journal of Vocational Behavior*, *66*, 26–44.

Bakker, A. B. (2008). The work-related flow inventory: Construction and initial validation of the WOLF. *Journal of Vocational Behavior*, *72*, 400–414.

Bakker, A., Golub, T. L., & Majdarijavec, M. (2017). Validation of the study-related flow inventory (WOLF-S). *Croatian Journal of Education, 19*(1), 147–173.

Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *Cell Biology Education, 15*(4), 1–7. https://doi.org/10.1187/cbe.16-04-0148

Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in developmental science: Situations, shortcomings, solutions, and standards. *Developmental Review, 33*(4), 357–370. https://doi.org/10.1016/j.dr.2013.08.003

Chen, L., Ye, Y. C., Chen, M. Y., & Tung, I. W. (2010). Alegría! Flow in leisure and life satisfaction: The mediating role of event satisfaction using data from an acrobatics show. *Social Indicators Research, 99*(2), 301–313.

Christensen, K. B., Engelhard, G., & Salzberger, T. (2012). Ask the experts: Rasch vs. factor analysis. *Rasch Measurement Transactions, 26*(3), 1373–1378.

Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. New York: HarperCollins.

Csikszentmihalyi, M. (2003). *Good business: Flow, leadership and the making of meaning*. New York: Viking.

Cunningham, T. J., Berkman, L. F., Gortmaker, S. L., Kiefe, C. I., Jacobs, D. R., Seeman, T. E., & Kawachi, I. (2011). Assessment of differential item functioning in the experiences of discrimination index: The coronary artery risk development in young adults (CARDIA) study. *American Journal of Epidemiology, 174*(11), 1266–1274. https://doi.org/10.1093/aje/kwr253

Demars, C. E. (2013). A comparison of confirmatory factor analysis and multidimensional Rasch models to investigate the dimensionality of test-taking motivation. *Journal of Applied Measurement, 14*(2), 179–196.

Ewing, M. T., Salzberger, T., & Sinkovics, R. R. (2009). Confirmatory factor analysis vs. Rasch approaches: Differences and measurement implications. *Rasch Measurement Transactions, 23*(1), 1194–1195.

Geyser, I., Geldenhuys, M., & Crous, F. (2015). The dimensionality of the Work Related Flow Inventory (WOLF): A South African study. *Journal of Psychology in Africa, 25*(4), 282–287. https://doi.org/10.1080/14330237.2015.1078084

Kasa, M., & Hassan, Z. (2015). The role of flow between burnout and organizational citizenship behavior (OCB) among hotel employees in Malaysia. *Procedia – Social and Behavioral, 211*, 199–206.

Kasa, M., & Hassan, Z. (2016). Flow experience and organizational citizenship behaviour among hotel employees: Moderating effect of socio-cultural factor. *Procedia - Social and Behavioral Sciences, 224*, 101–108. doi: 10.1016/j.sbspro.2016.05.409

Kasa, M., & Hassan, Z. (2017). Burnout dimensions with work-family conflict among hotel employees: Flow experience as mediating role. *Journal of Management & Marketing Review, 2*, 1–7.

Linacre, J. M. (2002). What do infit and outfit, mean-square and standardized mean. *Rasch Measurement Transactions, 16*(2), 878.

Linacre, J. M. (2012). *A user's guide to Winsteps Ministeps Rasch-model computer programs [version 3.74. 0]*. Chicago IL: Winstep.com.

Moneta, G. (2004). The flow model of intrinsic motivation in Chinese: Cultural and personal moderators. *Journal of Happiness Studies, 5*, 181–217.

Meyer, J. P. (2014). *Applied measurement with JMetrik*. New York: Routledge.

Paulus, I. S. (2012). Model konseptual pemanfaataan teori flow dalam e-learning [Conceptual model of Flow theory benefits in e-learning]. *Jurnal Nasional Pendidikan Teknik Informatika, 1*(1), 24–28.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Seligman, M., Csikszentmihalyi, M., Fredrikson, B. L., Wateman, A., Emmons, R. (2011). *Cross cultural advancement in positive psychology*. London and New York: Springer Drodrecht Heidelberg.

Smith, A. B., Wright, P., Selby, P. J., & Velikova, G. (2007). A Rasch and factor analysis of the functional assessment of cancer therapy-general (FACT-G). *Health and Quality of Life Outcomes, 5*. https://doi.org/10.1186/1477-7525-5-19

Soltani, A., Roslan, S., Abdullah, M. C., & Jan, C. C. (2011). Facilitating flow experience in people with intellectual disability using a music intervention program. *International Journal of Psychological Studies, 3*(2), 54–63.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah, NJ: Erlbaum.

Wolfe, E. W., & Singh, K. (2011). A comparison of structural equation and multidimensional Rasch modeling approaches to confirmatory factor analysis. *Journal of Applied Measurement, 12*(3), 212–221.