(ISSN: 2602-4047)

Uluman Mert, M. (2023). Flourishing Scale: Evaluation of Psychometric Properties in Line with the Rasch Model, *International Journal of Eurasian Education and Culture , 8(23), 2407-2425.*

DOI: http://dx.doi.org/10.35826/ijoecc.790

Article Type: Research Article

FLOURISHING SCALE: EVALUATION OF PSYCHOMETRIC PROPERTIES IN LINE WITH THE RASCH MODEL

Müge ULUMAN MERT

Assistant Professor, Marmara University, İstanbul, Türkiye, mugeulumann@gmail.com

ORCID: 0000-0003-4155-3114

Received: 15.01.2023 Accepted: 18.09.2023 Published: 01.10.2023

ABSTRACT

One of the important issues emphasized within the scope of positive psychology is the concept of flourishing. Flourishing, whose relationship with many different variables has been demonstrated in the literature, is also frequently used in psychology and educational sciences research. Based on these reasons, it is considered important to examine the structure of flourishing and the psychometric properties of the scales measuring this structure. The aim of this study is to determine the psychometric properties of the short form of flourishing scale within the scope of the Rasch model and to reveal whether differential item function according to the gender variable. The study group of the research consists of 576 university students studying at a public university in Istanbul. The Short Form of the Flourishing Scale, which is widely used in social sciences and consists of eight items and a single dimension, was used as a data collection tool. The analysis of the data obtained was made according to the Rasch Partial Credit Model. First of all, the assumptions of the Rasch model, namely unidimensionality and local independence, were tested. It was later revealed that all items were compatible with the model. In addition, it was found that the seven-point scale used in the scale worked effectively and the observed and expected values in the item characteristic curves overlapped. Finally, according to the Likelihood Ratio results, the fourth item showed non-uniform, the sixth and eighth items showed uniform negligible Differential Item Functioning. The scale was evaluated within the scope of the Rasch model, and the relevant scale can be evaluated within the scope of Generalizability Theory in different studies.

Keywords: Flourishing, Rasch, partial credit model, differential item functioning.

INTRODUCTION

Psychological well-being has been promoted as a psychological need and is considered an important phenomenon for human health and development (Hascher, 2010). It is referred to in various ways as psychological flourishing, life satisfaction, happiness, or finding meaning in life (Howell et al., 2016). One of the important issues emphasized within the scope of positive psychology is the concept of flourishing (Demirbaş, 2010; Miller & Foster, 2010; Seligman & Csikszentmihalyi, 2000). Flourishing can best be defined as leading a life that is oriented towards being healthy, in which the mind, spirit and body are united, integrated, having individual goals and a more meaningful life, socially, personally and ecologically functional (Myers et al., 2008). According to another definition, it can be expressed as the best health and welfare state that each individual has the power to achieve (Ivey et al., 2013). According to the theory on which flourishing is based, it states that when individuals focus not only on the problems but also on the developmental powers that exist within them (Myers et al., 2010). A low level of flourishing can provide information about which needs of individuals are not met and which developmental strengths they should focus on (Gündoğdu & Yavuzer, 2012). A high level of flourishing, on the other hand, can provide information about individuals' high motivation in solving the problems they encounter and their tendency to produce creative solutions (Aral, 2020). Therefore, it is important to measure the flourishing of individuals.

In the literature, it is seen that the relationship between flourishing and many different variables and concepts (altruism, psychological resilience, quality of life, etc.) has been modeled and studied. There are studies examining whether the flourishing of individuals differs depending on whether they are male or female (Doğan & Yıldırım, 2006; Gürgân, 2014; Haymana & Kolburan, 2019; Korkut-Owen et al., 2017; Laska et al., 2009;), although studies investigating the change in flourishing according to the age variable (Aral, 2020; Owen & Çelik, 2018; Korkut-Owen et al., 2017; Kasapoğlu, 2014) are more common, socio-economic level (Bahar & Başıbüyük, 2019), social environment (Kasapoğlu, 2013); It is seen that the variables of the country of citizenship (Öngöre, 2018) are also studied. According to the literature, flourishing and therefore the flourishing scale can be stated as a frequently studied subject and a measurement tool used.

There are many measurement tools that measure flourishing with social, emotional, physical, intellectual, spiritual, psychosocial, situational and environmental dimensions. Among these tools, Hettler's (1980) "The Wellness Inventory of the Life Assessment Questionnaire" with five dimensions and 100 items, Adams et al.'s (1997) six-dimensional and 36-item "The Perceived Wellness Survey", Renger et al.'s (2000)) "Optimal Living Profile" with six dimensions and 135 items, "Wellness Evaluation of Life Inventory" by Myers et al. (1998) with five dimensions and 114 items, "Wellness Evaluation of Life Inventory" by Travis (1981) with 12 dimensions and 120 items, Diener and colleagues (2010) "Short Form of Flourishing Scale (FL)", which is one-dimensional and contains 8 items, can be given as an example. When all these measurement tools are evaluated together, the fact that the Short Form of Flourish Scale addresses contexts such as respect and optimism that are not addressed

in other well-being scales, is unidimensional, and contains a small number of items can provide a significant advantage in terms of the usefulness and applicability of the scale.

When the literature is examined, it is seen that the Flourishing Scale has been adapted to many different languages, including Chinese, French, Russian, Lithuanian, Colombian, Spanish and Turkish. It can be stated that these studies were carried out with adult groups, the sample sizes were sufficient and varied between 320 and 1255, the genders were selected balanced, but this could not be achieved in age groups, and analyzes were made according to Classical Test Theory (CTT) in all of them, except for the study on adaptation to French culture. Considering the results of these adaptation studies based on CTT, a unidimensional structure was achieved, the explained variance varied between 37% (Russia) and 70% (Colombia), reliability in terms of internal consistency was in the range of .83-.93, and item-total correlations were .41. and .76, and the fit indices are sufficient (CFI in the range of .94-.98; TLI in the range of .92-.96; CI >.90; RMR and RMSEA values are lower than .08). In this context, it has been reported that the adapted scales can be used as a valid and reliable measurement tool (Checa et al., 2017; Didino et al., 2019; Doğan et al., 2012; Durak & Durak, 2019; Martin-Carnobell et al., 2021; Sadauska & Kolesovs, 2021; Salama-Younes, 2016; Tang et al., 2014; Telef, 2013).

In this research, the short form of flourishing scale, which was developed by Diener and colleagues in 2010 and adapted into Turkish by Fidan & Usta (2013), was used. In line with the examinations carried out, no research could be found in which the Rasch model was studied and its psychometric properties were determined, except for a study in which the well-being scale was adapted into French. Therefore, it is considered very important to determine the psychometric properties of this scale by Rash analysis.

There are different theories used in determining psychometric properties. The most frequently used of these theories in the literature is the CTT. Another theory that is relatively new compared to classical test theory is the Rasch measurement model. The Rasch model has many advantages in terms of scale development and adaptation, and researchers have stated that the use of the Rasch model in Likert-type scales helps to overcome various limitations regarding CTT-based methods (Elhan & Atakurt, 2005; İlhan & Güler, 2018). In the analyzes carried out according to CTT, the differences between the categories of Likert-type items are assumed to be equal. However, the items are actually at the order level. Therefore, analyzes made by taking the total score from these items, which are considered to be measured at equal interval level in the CTT, should be interpreted carefully and may give biased results. In addition, in CTT, the difficulty levels of all items are assumed to be equal and differences between difficulty levels are ignored (Anshel et al., 2009; Brinthaupt & Kang, 2014). Contrary to all these situations, the Rasch model; By performing a logit transformation, the difficulty level in transitions between categories is calculated, thus taking into account the unequal categories (Wright & Masters, 1982). In analyzes based on CTT, reliability estimation cannot be made at each item level. However, in Rasch model analyses, reliability estimation can be made at the item level for each individual (DeMars, 2010; Güler et al., 2017; Taşdelen et al., 2015). In addition to all these, the advantages of the Rasch model include ensuring parameter invariance (item parameters independent of the group and ability estimation independent of the item

sample), less need for norm studies, being useful in determining item bias, and calculating reliability estimates for each individual and ability (Boone, 2016; DeMars, 2010; Embretson & Reise, 2000; Engelhard, 2013; Hambleton & Swaminathan, 1985; Price, 2017).

Rasch Partial Credit Model (PCM) was used in this study. This model was developed for situations where the response categories of the item are consecutive, but the distances between these response categories are not equal (Koch and Dodd, 1989). "Step difficulty (category intersection parameter)" defined in the Rasch PCM model refers to the probability of an individual responding to one category compared to the other (Garrett, 2009). In the Rasch model, it is assumed that the discrimination of all items is equal. Therefore, there is no discrimination parameter "a" in the model. The following formula is used in the background of the PCM model, where the individual's ability estimate is θ and step difficulty is β:

$$P_{ijx} = \frac{exp \sum_{k=0}^{x}(\theta_i - \beta_{jk})}{\sum_{k=0}^{m} exp \sum_{t=0}^{k}(\theta_i - \beta_{jt})}$$

In scale development or adaptation studies on the psychological flourishing structure, it is seen that the use of the Rasch model is very low, despite all its advantages, and studies are mostly based on the CTT. In this context, the general aim of this study is to determine the pychometrics properties of the short form of flourishing scale within the scope of the Rasch model and to reveal whether it shows a differential item function according to the gender variable. In line with this general purpose, the sub-objectives of the research are given below.

- What are the psychometric properties of the short form of flourishing scale in line with the Rasch model?

- Do the results obtained from the short form of flourishing scale show differential item function according to the gender variable?

METHOD

The aim of this study is to assess the psychometric properties of the Flourishing Scale in line with the Rasch Partial Credit model. Accordingly, the sub-headings of the study group, data collection tool and data analysis are given below.

Study Group

The study group of the research consists of 576 university students studying at a public university in Istanbul in the fall semester of 2018-2019. Of the university students, 78.5% were female (N=452) and 21.5% were male (N=124). While deciding on the number of people to be included in the study group, the minimum sample size (500) recommended for the studies conducted within the scope of the Rasch model was taken into account. When working with a number smaller than this sample size, it will be difficult to ensure the invariance of the item-ability parameters and also parameter estimates will be inaccurate (de Ayala, 2009; DeMars, 2010).

Data Collection Tool

The short form of the Flourishing Scale was used in the study. The short form of the Flourishing Scale was developed by Diener and colleagues in 2010 and adapted into Turkish by Fidan & Usta (2013). This scale aims to measure individuals' perceptions of their own Flourishing Scale. The scale, which consists of a total of eight items, can be answered in a 7-point likert type ranging from strongly disagree to completely agree. There is no weighting for the scale items and the total scores of the individuals from the scale can be calculated by summing the responses given to the items. Therefore, the lowest score obtained from the scale is eight, and the highest score is 56. There is no item in the scale that requires reverse scoring. High total scores obtained from the scale indicate that the individual's level of Flourishing is high, and low total scores obtained indicate that the individual's level of Flourishing is low. As in its original form, Fidan & Usta (2013) also found out that the scale has a one-dimensional structure. According to the results obtained from this study, the factor loading values of the scale are between .61 and .77. The Cronbach's alpha coefficient, which is an indicator of the reliability of the scale in terms of internal consistency, was found to be .87. The goodness of fit indices obtained according to the results of the confirmatory factor analysis carried out to determine the construct validity are as follows: x 2 =48.80, sd=18, p=0.0001, RMSEA=.066, NFI=.97, CFI=.98, IFI =.98, RFI=.96, GFI=.97, and SRMR=.038.

Data Analysis

First of all, the assumptions needed to perform Rasch analysis; unidimensionality, local independence, and model-data fit (DeMars, 2010; Glas & Verhelst, 1995) were tested. When total scores are not unidimensional they are technically invalid (Smith 2002). Principal Axis Factor Analysis was used for the unidimensionality assumption (de Ayala, 2009), which reveals whether the observed variables are a function of only a single latent variable. Principal Axis Factor Analysis is considered one of the best methods to understand the nature of the structure behind the items (Costello & Osborne, 2005; Tennant & Pallant 2006). The local independence assumption implies that for a given skill level the items are unrelated. For the local independence assumption also called as response dependency, the main scale, there should be no leftover patterns in the residuals. A violation of the assumption of local independence of substances can be found in two ways: response dependence and multidimensionality (Hagell & Nilsson, 2009). There are studies indicating that there is no problem for this assumption if the unidimensionality assumption is met (Embretson & Reise, 2000; Lord, 1980; Morizot et al., 2007; Wright, 1996). However, even if the unidimensionality assumption is met, the local independence assumption should be reviewed (DeMars 2010; Reise, 1990). Therefore, in the study, the local independence assumption was examined using Yen's 3rd Quarter statistic by applying .20 criteria (Chen & Thissen, 1997; Christensen et al., 2017) with considering the .30 criterion (Røe, et al., 2014). AIC, BIC, CAIC, χ^2 value and significance level were used for model comparison. Item fit statistics were determined using infit and outfit values. And also fit statistics were used to identify items that misfit. From these statistics, infit value is more sensitive to the individual's responses to items with similar difficulty levels; The outfit value is more sensitive to the unexpected responses of the individual to the more difficult or simpler items. Infit pays attention to the

difference between observed and expected responses for items with a level of difficulty close to the person's ability level. outfit includes variances for all items, regardless of how far the item's difficulty is from one's ability (Smith & Smith, 2004). Although both statistics take values ranging from 0 to ∞, a value of 1.00 indicates perfect fit (Wright & Linacre, 1994). However, a value between 0.50 and 1.50 can be described as an indicator of item-model fit (Linacre, 2014). The category statistics in the Rasch analysis outputs and whether the response categories for each item are in order were examined. Item characteristic curves were obtained for each item in the scale. For model fit, the observed and expected characteristic curves are expected to overlap (Kaplan & Saccuzzo, 1989). In order to determine reliability, the Person Reliability value, also known as Person Separation Reliability, was used within the scope of the Rasch model. It can be stated that the closer the Person Reliability value obtained within the scope of the research is to 1.00, the higher the reliability. As with the Cronbach alpha internal consistency coefficient, it is recommended in the literature to take .70 as a criterion (Boone et al., 2014; Walker et al., 2012; Wright, & Stone, 1979).

After the Rasch model estimations were performed, it was determined whether the scale items showed differential item function (DIF). DIF is a statistic that gives the probability of individuals in different groups with the same ability levels to respond differently to the same item for the relevant variable to be measured (Holland & Wainer, 2012; Kamata & Vaughn, 2004; Roever, 2005; Zumbo, 1999). In this research, two types of DIF (uniform and non-uniform) coefficients are reported. Uniform DIF is where the group shows a consistent systematic variation in responses to an item across the entire range of the trait being measured. Non-uniform DIF and occurs when there is no uniformity in the differences between groups (Teresi et al., 2000). Within the scope of Rasch model and multi-category scoring, the likelihood ratio method was used to determine DIF according to gender in line with the purpose of the research. In the context of this method, the G2 value was obtained by taking the logarithm of the Likelihood Ratio and reported in the table in findings Chi-Square with the degrees of freedom. The significance of the G2 value indicates that the DIF is seen, while the value itself gives information about the size of the DIF. The ranges for this value and their meanings are as follows: $3.84 < G2 < 9.4$, no DIF or negligible; $9.4 \leq G2 < 41.9$ middle level of DIF; $G2 \geq 41.9$ high level of DIF (Drasgow et al., 2018; Greer, 2004; Thissen, 2001). With these values, as a result of the DIF analysis, the items are labeled in categories A (insignificant/nonsignificant DIF), B (medium DIF), and C (high DIF) (Narayanon & Swaminathan, 1996; Zieky, 1993). In the research, the above values were given together and the findings were interpreted. Jamovi 2.3.13 program was used for all analyzes carried out within the scope of the research.

FINDINGS

Before the analysis of the research, the assumptions of the Partial Credit Rasch Model were tested. Firstly, Principal-Axis Factor Analysis was carried out to determine whether the unidimensionality assumption was met. In this context, the obtained KMO value and Barlett Sphericity Test results were examined. The KMO value was found to be .90, and the Barlett's Test of Sphericity result was significant ($\chi 2 = 1069{,}242$, $sd = 28$) and it was determined that the data were suitable for factor analysis. As a result of Principal-Axis Factor Analysis, it was

found that the data had a one-dimensional structure (Figure 1) and factor load values varied between .61 and .82. The variance explained by this structure was obtained as 50.46%.

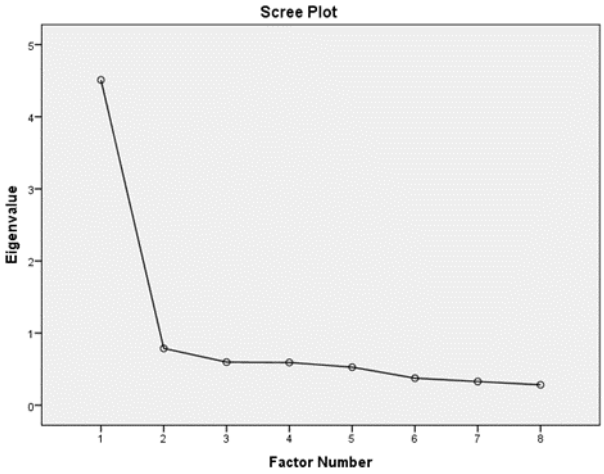

Figure 1. Scree plot for scale

It can be stated that when the unidimensionality assumption is met, the local independence assumption can also be considered to be met (Embretson & Reise, 2000). However, Yen's Q3 statistics, another indicator of the local independence assumption, were examined. Parameter estimations were made, residual values were found for each item, residual matrix was created and correlations between them were examined. The Q3 Correlation Matrix is presented in Table 1 below.

Table 1. Q3 Correlation matrix

	İ1	İ2	İ3	İ4	İ5	İ6	İ7	İ8
İ1	—							
İ2	-0.032	—						
İ3	0.163	-0.014	—					
İ4	-0.238	0.023	-0.005	—				
İ5	-0.193	-0.108	-0.117	-0.194	—			
İ6	-0.178	-0.046	-0.293	-0.084	-0.047	—		
İ7	-0.118	-0.227	-0.295	-0.296	-0.223	0.003	—	
İ8	-0.142	-0.120	-0.270	-0.173	-0.098	0.026	-0.058	—

Yen's Q3 statistics being over 0.30 can be considered as a violation of local independence. In line with this information, when the values in the Q3 correlation matrix in Table 1 are examined, it is seen that the Q3 correlations between the item pairs are less than .30 in absolute value. Considering the obtained values and this criterion, it is seen that the assumption of local independence is met. After Rasch analysis assumptions were

tested, it was statistically compared which model (Partial Credit Model (PCM), Rating Scale Model (RSM)) fit the data better. The results obtained are given in Table 2.

Table 2. Model comparison results

Model	AIC	BIC	CAIC	Log-Lik	χ^2	*df*	*p*
PCM	6631	6744	6758	-3267	131	35	< .001
RSM	6692	6811	6860	-3332			

In Table 2, which contains the statistical comparisons of the models, AIC, BIC, CAIC, log-Likelihood values, χ2 statistics and the p value of this value are given. The fact that the p value of the χ2 statistic is less than .05 can be interpreted as PCM being a more appropriate model for the scale items. However, it is known that AIC, BIC and CAIC statistics are better in model selection (Kankaras et al., 2010; Nylund et al., 2007). If these values of the models are smaller, it can be stated that the relevant model is more suitable. In this context, it can be stated that PCM is a more suitable model. Infit and Outfit values obtained within the scope of Rasch PCM for item-model fit are given in Table 3.

Table 3. Item statistics of the model

	Item mean	Measure	S.E.Measure	Infit	Outfit	Point biserial
İ1	4.10	0.527	0.0636	0.803	0.837	0.759
İ2	4.37	-0.352	0.0655	0.712	0.699	0.760
İ3	4.23	0.323	0.0590	0.972	0.981	0.712
İ4	4.23	-0.221	0.0657	1.164	1.183	0.663
İ5	4.22	0.319	0.0658	1.115	1.170	0.666
İ6	4.58	-0.628	0.0735	0.750	0.727	0.734
İ7	3.77	0.888	0.0550	1.432	1.447	0.668
İ8	4.78	-0.855	0.0763	1.063	1.000	0.633

Infit = Information-weighted mean square statistic

Outfit = Outlier-sensitive means square statistic

Infit and outfit values given in Table 3 are considered as item fit statistics. Both statistics are expected to be in the range of 0.50 to 1.50. When these values are examined, it is seen that Infit values vary between .71 and 1.43, while Outfit values vary between .70 and 1.45. In this context, it was determined that all items in the scale were compatible with the model. The point biserial correlation coefficient reported in the outcomes of the Rasch analysis gives information regarding whether all the components of any surface work in the same direction (Bond et al., 2007; Linacre, 2014). And also the coefficient refers to the correlation between the observations in the data and the measurements of the participants. A value of .30 can be considered as a criterion for the point biserial correlation coefficient, which is also considered as an item discrimination index (Linacre, 2021; Walter et al., 1998). The obtained correlation coefficient values are above the criterion .30, between .63 and .76. The threshold values for the response categories of the items are given in Table 4 below.

Table 4. Thresholds of the partial credit model

	Threshold 1	Threshold 2	Threshold 3	Threshold 4	Threshold 5	Threshold 6
İ1	-2.33	-2.22	-1.588	-0.738	0.2726	1.347
İ2	-12.00	-2.86	-1.785	-0.877	-0.0570	0.958
İ3	-2.99	-1.90	-1.341	-0.720	0.0182	0.758
İ4	-12.00	-2.78	-1.716	-0.832	0.2487	1.046
İ5	-2.90	-2.18	-1.790	-0.910	0.1643	1.331
İ6	-12.00	-3.02	-2.213	-1.539	-0.2299	1.104
İ7	-1.84	-1.39	-0.982	-0.387	0.3615	1.097
İ8	-12.00	-3.07	-2.555	-1.577	-0.7245	0.981

One of the advantages of the Rasch model is to determine how well the response categories work through information on the threshold parameters (Linacre, 2014). When the threshold parameters in Table 5 are examined, it is seen that they increase from the lowest to the highest. According to this result, it can be stated that the seven-point rating in the scale works effectively. The item characteristic curves of eight items are given in Figure 2.

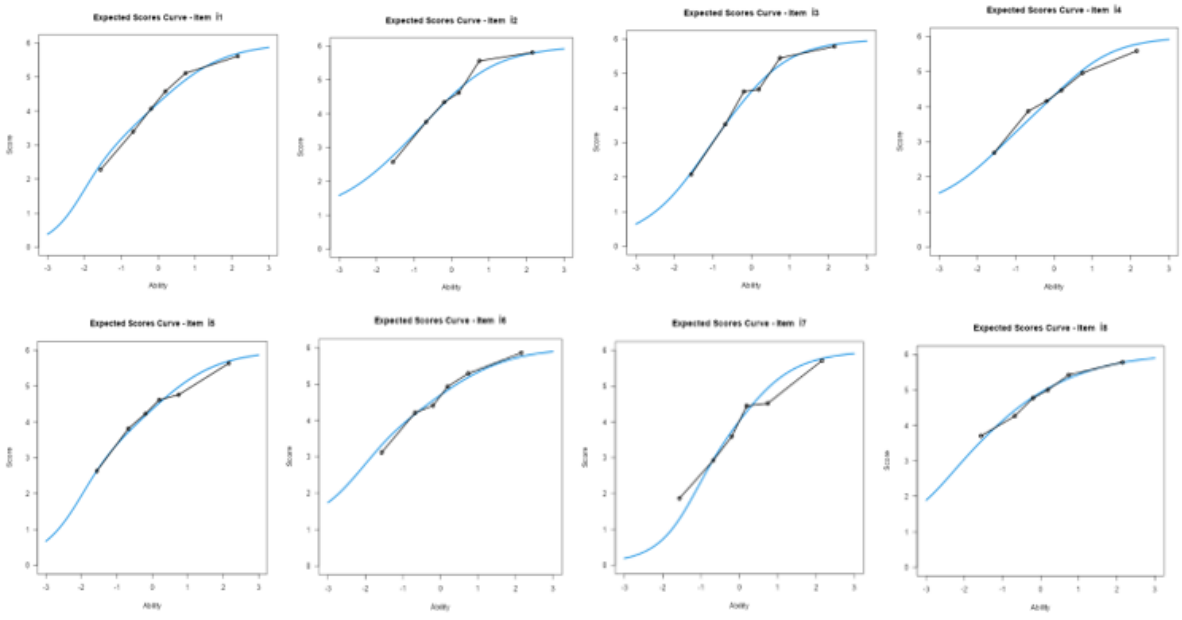

Figure 2. Item characteristic curve

In Figure 2, the black line appearing on the Item Characteristic Curves shows the responses of the individuals to the relevant item according to the data. The blue line is a theoretical item characteristic curve obtained by using the b parameter of the relevant item. When the item characteristic curves of each item are examined, it can be stated that the observed probabilities are compatible with the expected probabilities. The item category for partial credit model is given in Figure 3.

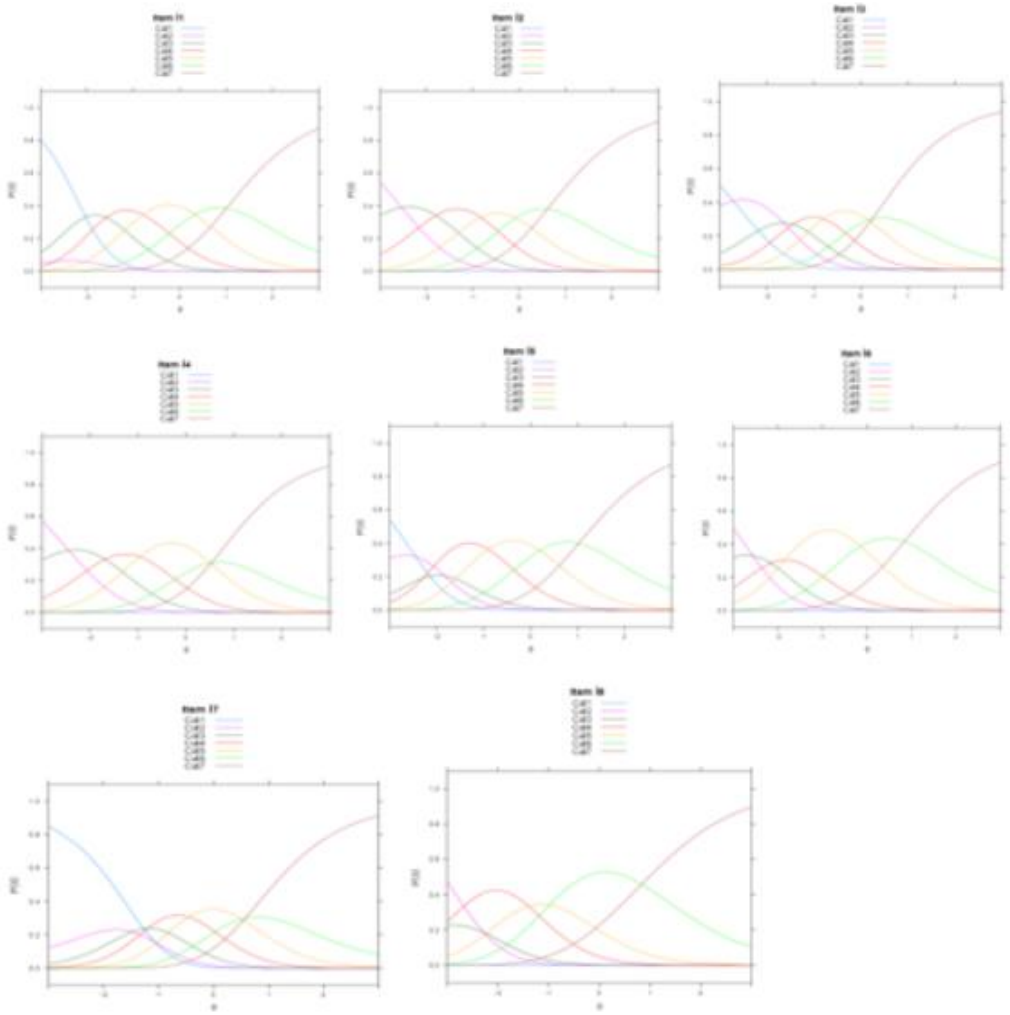

Figure 3. Item category for PCM

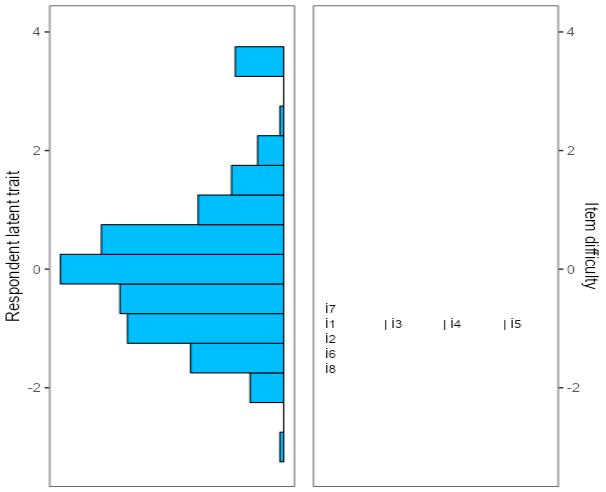

Figure 4. Wright Map

First of all, when the item category for PCM figures given above are examined, it can be said that as the skill levels on the horizontal axes for each item increase, the responses to the item have a sequential transition from 'I totally disagree' to 'I totally agree'. Figure 4 shows Wright Maps. This map is also called person-item maps and provides information about the distribution of individual responses (left side of the graph) and the distribution of item difficulties (right side of the graph). When the graph is examined, it can be stated that the 1st, 3rd, 4th and 5th items have similar difficulty levels. Person reliability value is considered as an indicator of reliability within the scope of the Rasch model. This value is considered equivalent to the Cronbach alpha coefficient, should be greater than 0.70 (Büyüköztürk et al., 2017) and is expected to take values greater than 0.80 (Fisher, 2007; Linacre, 2014) can be mentioned. The Person Reliability value obtained within the scope of the research is .84 and it can be stated that the reliability of the scale is high. Finally, it was examined whether the items in the scale showed DIF. In this context, the findings obtained in line with the Likelihood Ratio method are given in Table 5.

Table 5. DIF results by likelihood ratio method

	Non-uniform		Uniform	
	G^2	*p*	G_2	*p*
Item 1	0.0150	0.902	0.0094	0.923
Item 2	0.7770	0.378	1.1813	0.277
Item 3	0.3412	0.559	0.0125	0.911
Item 4	5.6853	0.017	0.0416	0.838
Item 5	0.4769	0.490	0.0501	0.823
Item 6	1.7704	0.989	6.8129	0.009
Item 7	0.7160	0.397	2.1315	0.144
Item 8	0.4457	0.504	6.1998	0 .011

When the values in Table 5 are examined, it can be stated that the G2 value of the fourth item within the scope of Non-uniform DIF is significant and shows DIF. However, it can be stated that the G2 value of this item (5.69) is in the range of 3.84 < G2 < 9.4 and therefore this item shows DIF at the no DIF or negligible level (A). Within the scope of Uniform DIF, it can be stated that the G2 value of the sixth and eighth items is significant and shows DIF. It is seen that the G2 value of the sixth (6.81) and eighth (6.20) items, similar to the fourth item, is in the range of (6.81) 3.84 < G2 < 9.4, and therefore this item shows (A) DIF at the no DIF or negligible level (Drasgow et al., 2018; Greer, 2004; Thissen, 2001). Therefore, while the fourth, sixth and eighth items show a negligible level of (A) DIF depending on gender, it can be stated that all three items works differently for reference and focus groups negligibly.

CONCLUSION and DISCUSSION

In this study, the psychometric properties of the Short Form of Flourishing Scale, which was developed by Diener and colleagues in 2010, and adapted into Turkish by Fidan and Usta in 2013, were tested within the scope of the Rasch Partial Credit Model (PCM). When the literature is examined, it can be stated that methods based on classical test theory (CTT) are generally used in the studies of adapting the scales to Turkish, as in the study of

Fidan and Usta (2013). However, there is a lot of research indicating that item response theory (IRT) has stronger features for scale development and adaptation studies and offers stronger predictions than CTT (Akyıldız & Şahin, 2017; Bulut, 2018; Cappelleri et al., 2008; Chao et al., 2019; Çelen & Aybek, 2013; Çüm & Gelbal, 2022; David et al., 2018; Güler & Gelbal, 2010; Köse, 2015; Özer & Özkan, 2014; Prieto et al., 2003; Sumintono, 2018; Van Zile-Tamsen, 2017). With this, no research providing IRT-based prediction of flourishing has been found in the Turkish literature. In line with these reasons, the presentation and interpretation of the psychometric properties of the IRT-based Rasch PCM of the short Form of Flourishing Scale adapted into Turkish by Fidan and Usta (2013) can be considered remarkable. First of all, the unidimensionality assumption, one of the assumptions of Rasch PCM, was tested using Principal Axis Factor Analysis. This analysis is frequently used by researchers for the relevant purpose (Brown, 2015; Chao et al., 2019; Piquero et al., 2000; Prieto et al., 2003). According to the Principal Axis Factor Analysis findings, a unidimensional structure was revealed with 50.46% explained variance. It is considered sufficient for the explained assumption to be 40% and above, especially for research conducted in the field of social sciences (Stevens, 1992). While the variance explained for the one-dimensional structure in the original form of the scale was 53%, it was determined as 47% in the study of Fidan & Usta (2013). The explained variance values obtained for other adaptation studies of the scale vary between 37% (Russia) and 70% (Colombia) (Checa et al., 2017; Didino et al., 2019; Doğan et al., 2012; Durak & Durak, 2019; Martin-Carnobell et al., 2021; Sadauska & Kolesovs, 2021; Salama-Younes, 2016; Tang et al., 2014; Telef, 2013). Salama-Younes, (2016), who conducted studies within the scope of the scale Rasch model, obtained this value as > 50%. For another assumption, the local independence assumption, Yen's 3rd Quarter statistic was used, which shows whether the answer to one item affects the other item. It can be stated that local independence is met in line with the correlations between the item pairs, thus providing reliable, valid and more accurate estimations for parameter estimations (Lee, 2004; Yen, 1993; Wainer & Thissen, 1996).

In line with the research findings, it is seen that infit values vary between .71 and 1.43, and outfit values vary between .70 and 1.45. Item difficulty levels and person ability levels in Rasch analysis as a result of the compatibility tests (infıt and outfıt), unexpected answers are determined (Bond et al., 2007). Additionally, while infit is an information-weighted fit statistic that is sensitive to the responses to test items, outfit is a fit statistics that is sensitive to outliers. Values for the fit statistics between 0.5 and 1.5 are considered appropriate for the measurement and both fit. The expected average value for the statistic is accepted as 1.0 (Hetherman, 2004; Linacre, 2002). In this context, the infit and outfit findings obtained show that all items in the scale are in the appropriate range for measurement and are in harmony. For this study, it was observed that threshold parameters increased from the lowest to the highest across response categories. Based on this finding, it can be determined that the seven -point grading system is effective.

Another indicator evaluated to determine the construct validity of the measurement tool is the results of the point biserial values of the items. According to the findings, it was seen that the point biserial values of the items varied between .63 and .76. High values obtained for the correlation coefficient mean that an item is able to

distinguish the abilities of respondents (Linacre, 2021). In this regard, since the values of point biserial correlation are greater than 0.30 (Bond et al., 2007), it was concluded that all items function in the same direction and should remain in the scale. On the other hand, the person reliability value obtained from the research was found to be .84. The results obtained from Rasch model person reliability analyzes are expected to have values greater than 0.80 (Bond et al., 2007; Fisher, 2007; Linacre, 2014). In this context, it can be concluded that the scale can be described as reliable. Fidan & Usta (2013), who adapted the scale to Turkish, found this value to be .83 in their study. For eight items, it was examined whether the scale showed DIF according to gender. According to the likelihood ratio statistical values, it was determined that the fourth, sixth and eighth items show a negligible level of (A) DIF depending on gender.

SUGGESTIONS

In this study, in line with the above results, it can be stated that the short form of flourishing scale gives valid and reliable results within the scope of Rasch PCM. However, it can be noted that their findings contain certain limitations. These include the fact that most of the individuals in the study group were women, the DIF was determined only by the likelihood ratio method, and secondary analyzes were not carried out with another method for the items showing DIF. Therefore, it is recommended that the research findings be evaluated within the framework of all limitations. Wıthın the scope of research, it can be suggested that a second DIF determination method can be used to determine the DIF and expert opinion can be taken to interpret item bias. However, the scale was evaluated within the scope of the Rasch model, and the relevant scale can be evaluated within the scope of Generalizability Theory in different studies.

ETHICAL TEXT

In this article, journal writing rules, publishing principles, research and publication ethics rules, and journal ethics rules have been followed. Responsibility for any violations that may arise regarding the article belongs to the authors. The data of this study were collected in the fall semester of 2018-2019.

Author(s) Contribution Rate: The author's contribution rate is 100%.

REFERENCES

Adams, T., Bezner, J., & Steinhardt, M. (1997). The conceptualization and measurement of perceived wellness: Integrating balance across and within dimensions. *American Journal of Health Promotion, 11*, 208–218.

Akyıldız, M., & Şahin, M. D. (2017). Açıköğretimde kullanılan sınavlardan Klasik Test Kuramına ve Madde Tepki Kuramına göre elde edilen yetenek ölçülerinin karşılaştırılması. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, *3*(4), 141-159.

Anshel, M.H., Weatherby, N.L., Kang, M., & Watson, T. (2009). Rasch calibration of a unidimensional perfectionism inventory for sport. *Psychology of Sport and Exercise*, *10*(1), 210-216. http://dx.doi.org/10.1016/j.psychsport.2008.07.006

Aral, S. (2020). *Egzersiz katılımcılarının özyeterlik, umut ve iyilik halinin incelenmesi* (Publication No. 645444) [*Y*üksek lisans tezi, İstanbul Gelişim Üniversitesi]. Ulusal Tez Merkezi.

Bahar, N. T. & Başıbüyük, H. H. (2019). Antalya 60+ Tazelenme Üniversitesi öğrencilerinin iyilik halinin değerlendirilmesi. *Ankara Sağlık Bilimleri Dergisi*, *8*(1), 160-173.

Bond, T. G., Fox, C. M., & Lacey, H. (2007). *Applying the Rasch model: Fundamental measurement. In the Social - Sciences (2nd)*. Routledge.

Boone, W.J., Staver, J.R., Yale, M.S. (2014). *Person Reliability, Item Reliability, and More. In: Rasch Analysis in the Human Sciences*. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-6857-4_10

Boone, WJ. (2016). Rasch analysis for instrument development: why, when, and how? *CBE-Life Sciences Education*, 15(rm4):1-7. https://doi.org/10.1187/cbe.16-04-0148

Brinthaupt, T.M., & Kang, M. (2014). Many-faceted rasch calibration: An example using the self-talk scale. *Assessment*, *21*(2) 241-249. http://dx.doi.org/10.1177/1073191112446653.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). The Guilford.

Bulut, G. (2018). Açık ve uzaktan öğrenmede şans başarısı: Klasik Test Kuramı (KTK) ve Madde Tepki Kurama (MTK) temelinde karşılaştırmalı bir analiz. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, *4*(1), 78-93.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2017). *Bilimsel araştırma yöntemleri*. Pegem Yay. Ankara.

Cappelleri, J. C., Lundy, J. 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

Chao, R. C. L., Green, K., Dugar, K., Paiko, L., Zhao, J., & Li, P. F. (2019). Rasch analysis of the Brief Symptom Inventory-18 with African American and White students. *Psychological Assessment*, *31*(1), 100. https://doi.org/10.1037/pas0000633

Checa, I., Perales, J. & Espejo, B. (2017). Spanish Validation of the Flourishing Scale in the General Population, *Current Psychology, 37*(4), 949-956, https://doi.org/10.1007/s12144-017-9581-0

Chen, W. and Thissen, D. 1997: Local dependence indices for item pairs distribution. *Educational and Behavioral Statistics* 22, 265–89. https://doi.org/10.3102/10769986022003265

Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q3: Identification of local dependence in the Rasch model using residual correlations. *Applied psychological measurement*, *41*(3), 178-194. 10.1177/0146621616677520

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, *10*(1), 7.

Çelen, Ü., & Aybek, E. C. (2013). Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Journal of Measurement and Evaluation in Education and Psychology*, *4*(2), 64-75.

Çüm, S. & Gelbal, S. (2022). Sato Test Kuramı'nın Klasik Test Kuramı ve Madde Tepki Kuramı ile Psikometrik Açıdan Karşılaştırılması. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 23(2), 1797-1829.

David, S. L., Hitchcock, J. H., Ragan, B., Brooks, G., & Starkey, C. (2018). Mixing interviews and Rasch modeling: Demonstrating a procedure used to develop an instrument that measures trust. *Journal of Mixed Methods Research*, *12*(1), 75-94. DOI: 10.1177/1558689815624586

de Ayala, R. J. (2009). *The theory and practice of ıtem response theory*. The Guilford Press.

DeMars, C. (2010). *Item response theory*. Oxford University Press.

Demirbaş, N. (2010). *Yaşamda anlam ve yılmazlık*. (Publication No. 265337) [*Y*üksek lisans tezi, Hacettepe Üniversitesi]. Ulusal Tez Merkezi.

Didino, D., Taran, E. A., Barysheva, G. A. & Casati, F. (2019). Psychometric evaluation of the Russian version of the flourishing scale in a sample of older adults living in Siberia. *Health and Quality of Life Outcomes*, *17*, 1-12. https://doi.org/10.1186/s12955-019-1100-6

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social İndicators Research*, *97*, 143-156.

Doğan, T., Yıldırım, İ. & Myers, J.E. (2012). Adaptation of the Wellness Evaluation of Lifestyle Scale to Turkish. *Eğitim Araştırmaları-Eurasian Journal of Educational Research, 46,* 23-40.

Drasgow, F., Nye, C. D., Stark, S., & Chernyshenko, O. S. (2018). Differential item and test functioning. In P. Irwing, T. Booth, & D. J. Hughes (Eds.), *The Wiley handbook of psychometric testing: A multidisciplinary reference on survey, scale and test development* (pp. 885–899). Wiley Blackwell. https://doi.org/10.1002/9781118489772.ch27

Durak, E.Ş. & Durak, M. (2019). Psychometric properties of the Turkish version of the Flourishing Scale and the Scale of Positive and Negative Experience. *Mental Health, Religion & Culture, 22*(10), 1021-1032. DOI: 10.1080/13674676.2019.1689548

Elhan, A. H., & Atakurt, Y. (2005). Ölçeklerin değerlendirilmesinde niçin Rasch analizi kullanılmalıdır?. *Ankara Üniversitesi Tıp Fakültesi Mecmuası*, *58*(1), 47-50.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists. Quality of Life Research*. Lawrence Erlbaum Associates.

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

Fidan, M., & Usta, F. (2013). İyilik hali ölçeğinin Türkçe formunun güvenirlik ve geçerliğinin incelenmesi. *Eğitim ve Öğretim Araştırmaları Dergisi*, *2*(3), 265-269.

Fisher, W. P. Jr. (2007). Rasch measurement transaction. transaction of the rasch measurement. *American Educational Research Association*. *21*(1), 1087–1096.

Garrett, P. (2009). *A Monte Carlo study ınvestigating missing data, differential item functioning and effect size* (Doctoral Dissertation). Georgia State University. doi: https://doi.org/10.57709/1060078

Glas, C. A. W., & Verhelst, N. D. (1995). *Testing the Rasch model*. In G. H. Fischer and I. W. Molenaar (Eds), Rasch models: Foundations, recent developments, and applications (pp. 69–96). New York: Springer-Verlag.

Greer, T. G. (2004). *Detection of differential item functioning (DIF) on the SATV: A comparison of four methods: Mantel-Haenszel, logistic regression, simultaneous item bias, and likelihood ratio test*. University of Houston.

Güler, N., İlhan, M., Güneyli, A., & Demir, S. (2017). An evaluation of the psychometric properties of three different forms of daly and miller's writing apprehension test through Rasch analysis. *Educational Sciences: Theory & Practice*, *17*(3), 721–744. http://dx.doi.org/10.12738/estp.2017.3.0051

Gürgân, U. (2014). Üniversite Öğrencilerinin Yılmazlık Ve İyilik Halinin Bazı Değişkenlere Göre İncelenmesi. *Education Sciences*, *9*(1), 19-35.

Hagell, P., & Nilsson, M. H. (2009). The 39-item Parkinson's Disease Questionnaire (PDQ-39): is it a unidimensional construct?. *Therapeutic Advances in Neurological Disorders*, *2*(4), 205-214. https://doi.org/10.1177/1756285609103726

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and appliccations*. Springer Science ve Business Media, LLC.

Hascher, T., (2010). *Wellbeing*. In: Peterson, Penelope; Baker, Eva; McGaw, Barry (eds.). International Encyclopedia of Education 6 (pp. 732-738). Elsevier 10.1016/B978-0-08-044894-7.00633-3

Haymana, P., & Kolburan, Ş. G. (2019). Kadınların iyilik halinin algılanan toplumsal cinsiyet rolleriyle ilişkisi. *Aydın İnsan ve Toplum Dergisi*, *5*(1), 29-53.

Hettler, B. (1980). Wellness promotion on a university campus: Family and community health. *Journal of Health Promotion and Maintenance*, 3, 77–95.

Holland, P. W., & Wainer, H. (2012). *Differential item functioning*. Routledge.

Howell, K. H., Coffey, J. K., Fosco, G. M., Kracke, K., Nelson, S. K., Rothman, E. F., & Grych, J. H. (2016). Seven reasons to invest in well-being. *Psychology of Violence*, 6(1), 8.

Ilhan, M., & Guler, N. (2018). A comparison of difficulty indices calculated for open-ended items according to classical test theory and many facet rasch model. *Eurasian journal of educational Research*, *18*(75), 99-114.

Ivey, A.E., Ivey, M.B., Myers, J.E., & Sweeney, T.J., (2013). *Gelişimsel psikolojik danışma ve terapi: Yaşam boyu iyilik halini arttırmak*. F. Korkut Owen, İngilizce ve Türkçe Edt, Ankara: Türk PDRDER Yayınları.

Kamata, A., & Vaughn, B. K. (2004). An introduction to Differential Item Functioning Analysis. *Learning Disabilities: A Contemporary Journal*, *2*(2), 49-69.

Kankaras, M., Moors, G., & Vermunt, J. K. (2010). *Testing for measurement invariance with latent class analysis*. In E. Davidov, P. Schmidt, & J. B. Billiet (Eds), Cross-Cultural analysis: Methods and applications. Routledge.

Kaplan RM, & Saccuzzo DP (1989). *Psychological testing: Principles, applications, and applications* (2nd ed.). Pacific Grove.

Kasapoğlu, F. (2013). *Üniversite öğrencilerinde iyilik hali ile özgecilik arasındaki ilişkinin incelenmesi*, (Publication No. 326664) [*Y*üksek lisans tezi, İnönü Üniversitesi]. Ulusal Tez Merkezi.

Kasapoğlu, F. (2014). İyilik hali ile özgecilik arasındaki ilişkinin incelenmesi, *Hikmet Yurdu*, 7(13), 271 - 288.

Koch, W. R. & Dodd, B. G. (1989). An investigation of procedures for computerized adaptive testing using partial credit scoring. *Applied Measurement in Education*, 2(4), 335-357.

Korkut-Owen, F., Demirbaş-Çelik, N., & Doğan, T. (2017). Üniversite Öğrencilerinde İyilik Halinin Yordayicisi Olarak Psikolojik Sağlamlik. *Elektronik Sosyal Bilimler Dergisi*, *16*(64-Ek Sayı), 1461-1479.

Köse, İ. A. (2015). Aşamalı tepki modeli ve klasik test kuramı altında elde edilen test ve madde parametrelerinin karşılaştırılması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *15*(2), 184-197.

Lee, Y. W. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q3 statistics in an EFL reading comprehension test. *Language Testing*, *21*(1), 74-100. https://doi.org/ 10.1191/0265532204lt260oa

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

Linacre, J. M. (2009). *A User's guide to Winsteps-ministep: Rasch-model computer programs*. Program Manual 3.68. 0. Chicago, IL.

Linacre, J.M. (2014). *A user's guide to FACETS Rasch-model computer programs*. http://www.winsteps.com/a/facets-manual.pdf

Linacre, J. M. (2021). *A User's guide to Winsteps-ministep: Rasch-model computer programs*. Program Manual 4.8.0. Chicago, IL

Lord, F. M. (1980). *Aplications of item response theory to practical testing problems*. Hillsdale, Lawrence Erlbaum Associates.

Martin-Carbonell, M., Espejo, B., Checa, I. & Fernández-Daza, M. (2021). Adaptation and Measurement Invariance by Gender of the Flourishing Scale in a Colombian Sample. *International Journal of Environmental Research and Public Health, 18*, 2664. https:// doi.org/ 10.3390/ijerph18052664

Miller, G.D. & Foster, L.T. (2010). *Critical synthesis of wellness literature*. http://silvrnetwork.net/sites/silvrnetwork.org/files/library/Critical_Synthesis%2520of%2520Wellness%2520Update.pdf adresinden alınmıştır.

Morizot, J., Ainsworth, A. T., ve Reise, S. P. (2007). Toward modern psychometrics. In R. W. Robins, R. C. Fraley, ve R. F. Krueger (Eds.), Handbook of research methods in personalty psychology. The Guilford Press

Myers, J. E., Sweeney, T. J., & Witmer, J. M. (1998). *The Wellness Evaluation of Lifestyle*. Greensboro.

Myers, J.E. Sweeney, T.J. & Witmer, J.M. (2000). The wheel of wellness counseling for wellness: A holistic model for treatment planning. *Journal of Counseling and Development*, 78, 251–266. https://doi.org/10.1002/j.1556-6676.2000.tb01906.x

Myers, J.E., Sweeney, T.J. & Witmer, J.M., (2010). *İyilik hali (wellness) çalışma kitabı*. (Çev.: F. Korkut Owen) Ankara.

Narayanon, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, *20*(3), 257-274. https://doi.org/10.1177/014662169602000306

Nylund, K. L., Asparouhov, T. & Muthn, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling*, 14(4), 535-569.

Owen, F. K., & Çelik, N. D. (2018). Yaşam boyu sağlıklı yaşam ve iyilik hali. *Psikiyatride Güncel Yaklaşımlar*, *10*(4), 440-453.

Öngöre, Ö. (2018). Sosyal Politikanın Bir Göstergesi Olarak Vatandaşların İyilik Halinin Ülkelere Göre Karşılaştırmalı Analizi. *Uşak Üniversitesi Sosyal Bilimler Dergisi*, *11*(C-IASOS Özel Sayısı), 1-10.

Özer Özkan, Y. (2014). Öğrenci başarılarının belirlenmesi sınavından klasik test kuramı, tek ve çok boyutlu madde tepki kuramı modelleri ile kestirilen başarı puanlarının karşılaştırılması. *International Journal of Human Sciences*, *11*(1), 20-44.

Piquero, A. R., MacIntosh, R., & Hickman, M. (2000). Does self-control affect survey response? Applying exploratory, confirmatory, and item response theory analysis to Grasmick et al.'s self-control scale. *Criminology*, *38*(3), 897-930. https://doi.org/ 10.1111/j.1745-9125.2000.tb00910.x

Price, L. R. (2017). *Psychometric methods: Theory and practice*. New York, NY: The Guilford Press.

Prieto, L., Alonso, J., & Lamarca, R. (2003). Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health and quality of life outcomes*, *1*, 1-13.

Reise, S.P. 1990: A comparison of item- and person-t methods of assessing model-data-t in IRT. *Applied Psychological Measurement.* 14, 127– 37. https://doi.org/10.1177/014662169001400202

Renger, R. F., Midyett, S. J., Mas, F. G., Erin, T. E., McDermott, H. M., Papenfuss, R. L., Eichling, et al. (2000). Optimal Living Profile: An inventory to assess health and wellness. *American Journal of Health Promotion, 24,* 403–412.

Røe, C., Damsgård, E., Fors, T., & Anke, A. (2014). Psychometric properties of the pain stages of change questionnaire as evaluated by Rasch analysis in patients with chronic musculoskeletal pain. *BMC Musculoskelet Disord, 15*(1), 95. https://dx.doi.org/10.1186/1471-2474-15-95.

Roever, C. (2005). That's not fair! Fairness, bias, and differential item functioning in language testing. *SLS Brownbag*, 9(15), 1-14.

Sadauska, I. & Kolesovs, A. (2021). Validity And Reliability of The Flourishing Scale: Latvian Version. *Society, Integration,Education Proceedings of the International Scientific Conference 3, May 28th-29th*, 171-180, https://doi.org/10.17770/sie2021vol7.6402

Salama-Younes, M. (2016). Should French masters athletes be flourishing? Psychometric studies *Journal of Psychological Sciences, 2*(3), 99–113.

Seligman, M.P., (2007). *Gerçek mutluluk*. Ankara.

Smith EV Jr & Smith RM. (2004). *Introduction to Rasch measurement*. JAM Press.

Smith, Jr E.V. (2002) Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement* 3: 205–231.

Sumintono, B. (2018). Rasch model measurements as tools in assesment for learning. In *1st International Conference on Education Innovation (ICEI 2017)* (pp. 38-42). Atlantis Press.

Tang, X., Duan, W. Wang, Z. & Liu, T. (2014). Psychometric Evaluation of the Simplified Chinese Version of Flourishing Scale. *Research on Social Work Practice*, 1-9. https://doi.org/10.1177/1049731514557832

Telef, B.B. (2013). Psikolojik İyi Oluş Ölçeği: Türkçeye Uyarlama, Geçerlik ve Güvenirlik Çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H. U. Journal of Education) 28*(3), 374-384.

Tennant, A. & Pallant, J. (2006) Unidimensionality matters. *Rasch Meas Trans* 20: 1048–1051.

Teresi JA, Kleinman M, Ocepek-Welikson K. (2000). Modern psychometric methods for detection of differential item functioning: application to cognitive assessment measures. *Stat Med*; 19:1651– 83.

Thissen, D. (2001). *Software for the computation of the statistics involved in item response theory likelihood-ratio tests for differential item functioning*. University of North Carolina at Chapel Hill.

Travis, J. W. (1981). *The Wellness Workbook*. Mill Valley.

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

Wainer, H. & Thissen, D. (1996). How is reliability related to the quality of test scores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice*. 15, 22–29.

Walker, E. R., Engelhard Jr, G., & Thompson, N. J. (2012). Using Rasch measurement theory to assess three depression scales among adults with epilepsy. *Seizure*, *21*(6), 437-443.

Walter, S.D., Eliasziw, M. & Donner, A. (1998). Sample size and optimal designs for reliability studies. *Stat. Med*., 17, 101–110.

Wright B. D. & Linacre JM (1994). Reasonable mean-square fit values. *Rasch Measure Trans*. 8: 370

Wright B. D. (1996). Local dependency, correlations and principal components. *Rasch Meas Trans.*10:509 –11

Wright, B. D. & Masters, G. (1982). *Rating scale analysis: Rasch measurement*. MESA Press.

Wright, B. D. & Stone, M.H. (1979). *Best lest design*. MESA Press.

Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. Web: http://psycnet.apa.org/psycinfo/1993-97193-017 adresinden 17 Mayıs 2016'da erişilmiştir.

Zumbo, B. D. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores.* Directorate of Human Resources Research and Evaluation, Department of National Defense.