(ISSN: 2602-4047)

Tunç, E.B. (2023). Assessing the psychometric properties of the Brief Resilience Scale: A Rasch modeling approach, *International Journal of Eurasian Education and Culture, 8(23), 2116-2137.*

DOI: http://dx.doi.org/10.35826/ijoecc.789

Article Type (Makale Türü): Research Article

ASSESSING THE PSYCHOMETRIC PROPERTIES OF THE BRIEF RESILIENCE SCALE: A RASCH MODELING APPROACH

Emine Burcu TUNÇ

Asst. Prof., Marmara University, Istanbul, Turkey, burcupehlivantunc@gmail.com

ORCID: 0000-0002-8225-9299

Received: 13.01.2023 *Accepted: 03.09.2023* *Published: 01.10.2023*

ABSTRACT

Research indicates that resilience is the concept that best explains individuals' efforts to heal the effects of negative experiences and recovery processes after difficulties. Hence, it is considered substantial to study the construct of resilience and the psychometric properties of the scales that measure this structure. The objective of this study is to evaluate the psychometric properties of the Brief Resilience Scale within the scope of the Rasch Partial Credit model. In this direction, the aim was to contribute to the literature with a research study that may be an example of the use of the Partial Credit Model in the assessment of the scales. The study group of the research consists of 608 individuals. As a data collection tool, The Brief Resilience Scale, which is widely used in social sciences and consists of six items and one dimension, was used. The analysis of the obtained data was performed based on the Rasch Partial Credit Model. After the unidimensionality and local independence assumptions of the Rasch model were tested, it was determined that all the items were compatible with the model. At the same time, it was seen that the five-point scale utilized in the scale worked effectively, and that the observed and expected values in the item characteristic curves overlapped. Based on the outcomes of the Likelihood Ratio, according to gender, the sixth item showed non-uniform, and the fifth item showed uniform negligible Differential Item Functioning. In line with all these findings, it has been demonstrated that the Brief Resilience Scale provides valid and reliable results within the scope of Rasch PCM.

Keywords: Resilience, Brief Resilience Scale, Rasch, Partial Credit Model, Differential Item Functioning

INTRODUCTION

The non-uniform life cycle presents positive and negative experiences together. Individuals' reactions to challenging experiences such as stress, trauma, and pandemics that affect their routine actions differ. Some individuals have chaotic reactions like depression, anxiety, burnout, and physical, emotional, or mental exhaustion. Other people do not perceive the negativity as a threat; they regard it as a driving force for a new beginning, career opportunity, or advantage that will contribute to their development (Abel & Sewell, 1999; Brown & Nagel, 2004; Coutu, 2002; Kobasa, 1979). The diversity of people's reactions to similar experiences attracts the attention of researchers. Studies conducted on what makes the difference led researchers to the concepts of stress-resistant, invulnerable, and resilienceer. Researchers contribute to the creation of the resilience paradigm by pointing out that resiliency is the concept that best explains individuals' efforts to heal the effects of negative experiences and recovery processes after challenges (Doğan, 2015; Masten & Reed, 2002; Neenan, 2009).

Resilience, a concept of Latin origin, has been researched in many different disciplines, such as the health sector, social sciences, educational sciences, organizational studies, war, and psychology. This wide field of study causes the concept to be defined differently (Day *et al*., 2011; Simmons, 2011). The ability of individuals to adapt to new situations under difficult life conditions, the positive adaptation process that the individual achieves in spite of intense stressful situations, the capacity developed in the face of traumatic events, not being sick or depressed due to stress, recovery from all threatening situations, finding life meaningful despite bad experiences, the state of not feeling like a victim in the face of negative experiences, the process of building a bridge to reach the hopeful days ahead despite the current adversities, and the capacity of the individual to continue to recover after becoming active in stressful times are among the definitions of resilience. (Bernshausen & Cunningham, 2001; Coutu, 2002; Day & Gu, 2014; Earvolino-Ramirez, 2007; Ee & Chang, 2010; Giroux, 2007; Kobasa, 1979; Masten *et al*., 1990). Henderson & Milstein (2003) suggest that resilience is a capacity inherent in every individual. Furthermore, they emphasize that the resilience development process is life itself, and individuals have to go on with their lives by overcoming all negativity. Smith *et al*. (2008) state that the focus is on the ability to get rid of stress, despite resilience being defined as resistance to disease, adaptation, and the healing process by other researchers. The emphasis on stress in Smith *et al*. (2008)'s definition is noteworthy. It is observed that stress is fed from many sources and arises from problems that occur at any unexpected moment, sometimes known and sometimes not even noticed. Globalization and communication technologies are also directing the stressful changes that occur anywhere in the world into our daily lives. Therefore, stress is counted as an integral part of 21st-century human life (Fimian, 1984; Maddi & Khoshaba, 2005). Even though definitions of resilience vary and highlight different dimensions, the general consensus is that it means going one step better by overcoming difficult conditions, stress, traumatic events, threatening situations, bad experiences, and all negativities.

Dedication, taking responsibility, the belief that one can control events (internal locus of control), seeing change as an opportunity for development (Kobasa & Puccetti, 1983); accepting the facts as they are, not denying them;

finding life meaningful even in bad situations; using the available resources in an original and creative way in the face of challenges (Coutu, 2002); establishing social and positive relationships; being autonomous; having a positive outlook on the future; high self-confidence, self-esteem and motivation (Henderson & Milstein, 2003); being socially competent; problem-solving skills, and being purposeful (Benard, 1991) are among the characteristics of individuals with high resilience. These qualities constitute the personal dimension of the individual's assets, which are expressed as protective factors that reduce the negative effects of a stressful situation and increase the possibility of obtaining positive outcomes (Masten & Reed, 2002). Resilience, which is not just a personality trait, may develop and diminish in the social environment (Day *et al*., 2011). A considerable number of individuals who experience stressful circumstances actively seek assistance from various sources, such as religion, experts, literature, or friends (Peres *et al*., 2005; Peres *et al*., 2007), in order to increase their resiliency towards stressful events. A study on stress reactions in the United States after the events of September 11 by Schuster *et al*. (2001) revealed that seeking solace in religious practices like prayer, participating in religious activities, or having spiritual experiences emerged as the second most common coping mechanism, with a reported prevalence of 90% following the act of engaging in conversations with people, which was reported by 98% of the participants. Therefore, protective factors also have an environmental dimension. Environmental protective factors, also referred to as social capital, point to the human interaction necessary for healthy improvement and relief from stress (Benard, 2004).

There are many measurement tools that measure resilience with different dimensions in various groups. Windle *et al*. (2011) suggested that three measurement tools had the best psychometric properties. These measuring tools are as follows: Connor–Davidson Resilience Scale (CDRISC; Connor & Davidson, 2003) with 25 items, the Resilience Scale for Adults (RSA; Friborg *et al*., 2003) with 37 items, and the Brief Resilience Scale (BRS; Smith *et al*., 2008) with 6 items. In this study, the psychometric properties of the Brief Resilience Scale, which was developed by Smith *et al*. (2008) and adapted by Doğan (2015), were reviewed. The scale was adapted by Doğan (2015) within the context of Classical Test Theory (CTT). It has been assessed within the scope of the Rasch model in this study. Examining the international literature, it is seen that the Rasch model is widely utilized in determining the psychometric properties of Likert-type scales because the Rasch model has many advantages from the point of scale development and adaptation. Researchers (Elhan & Atakurt, 2005; İlhan & Güler, 2018) suggest that using the Rasch model in Likert-type scales helps overcome various limitations of CTT-based methods.

In Classical Test Theory (CTT), item parameters are affected by the characteristics of individuals. When the same items are applied to individuals in different groups, different item parameters can be obtained, so it is seen that the obtained item parameters are group dependent. However, as in all models within the scope of Item Response Theory (IRT), ability levels of individuals and item parameters are located along a common axis in the Rasch model. Individuals' ability levels are estimated independently of the item sample in the measurement tool, and item parameters can be calculated independently of the ability levels of the individuals in the test group (Boone, 2016; DeMars, 2010; Embretson & Reise, 2000; Engelhard, 2013; Hambleton & Swaminathan, 1985; Price, 2017). Likert-type scale results are at the ordering level. However, in CTT, the total score is taken as equal intervals, and parametric statistics are used. This may lead to biased results (Brinthaupt & Kang, 2014). In CTT, ability estimations of individuals are obtained by summing their responses to scale items. It is assumed that the difficulty levels of all the items in the scale are similar, and this causes the possible differences between the difficulty levels of the scale items to be ignored when calculating the ability estimation (Anshel *et al*., 2009). The Rasch model, on the other hand, helps to overcome these limitations by transforming the data in the ranking scale into an evenly spaced logit scale (Wright & Masters, 1982). At the same time, standard analysis methods, such as summing the scores for the answers given to the items, perform the analysis by regarding the intervals between the options as equal. In reality, the intervals between the options are not always equal (Elhan & Atakurt).

In CTT, it is accepted that the variability in the individual is caused by individual differences, and no information can be obtained on the reliability level of the scale items that would show to what extent they can distinguish between individuals (Taşdelen Teker *et al*., 2015). However, the reliability coefficient of each individual is reported in the Rasch model. Therefore, it is possible to ascertain the level of reliability in distinguishing between people within the group to which the scale is administered (Güler *et al*., 2017). Missing data is also not a problem in Rasch analysis because Rasch analysis only processes one observation at a time. That is, it calculates the expected value separately for each observation. Missing data is skipped during this calculation process. A person's ability level is calculated by comparing the sum of the observed values with the sum of the expected values for each person in the study. Similarly, item difficulty levels are calculated by comparing the sum of the observed values with the sum of the expected values for each item. Since this addition is done only on the data of known value, no correction or adjustment is required for missing data (Elhan & Atakurt).

Rasch Partial Credit Model (PCM) was used in this study. It was developed by PCM Masters in 1982 and is an extension of the Rasch model developed for two-category items. This model was developed for situations where it is important to give partial points in the case of completing different stages in the analysis process or when the distances between the response categories in Likert-type items differ from item to item. One of the important features of the model is that it is possible to score people with moderate θ levels using this model (Koch & Dodd, 1989).

In PCM, the individual parameter θ and the item parameter β are available. Masters defines β parameters as “step difficulty”. The reason it is defined as step difficulty is that after the individual successfully completes one step, s/he moves on to the next step. The item step difficulty parameter is also called the category intersection parameter. As a result, the step difficulty parameter is defined as the amount of difficulty involved in selecting another response category from one response category. In PCM, there is one missing step difficulty parameter from the item category number. For example, there are three-step difficulty parameters for a four-category item (Garrett, 2009). As in all Rasch models, items are assumed to be equally distinctive. Therefore, there is no item discrimination parameter in the model. PCM assumes that the steps are consecutive. However, this does not mean that every step is of equal difficulty; that is, it cannot be deduced that the last step is the hardest and the first step is the easiest (Koch & Dodd, 1989).

There are scale development and adaptation studies with different study groups within the scope of the Rasch model on resilience (Alavi *et al*., 2020; Jefferies *et al*., 2019; Kim *et al*., 2022; Li *et al*., 2019; Ricketts *et al*., 2015; Yaşar & Aybek, 2019). In this study, the psychometric properties of The Brief Resilience Scale, developed by Smith *et al*. (2008), adapted by Doğan (2015), and widely used in both education and psychology, were tested within the scope of Rasch PCM. The Brief Resilience scale has also been adapted in many cultures; Spanish (Rodríguez-Rey *et al*., 2015), Portuguese (Coelho *et al*., 2016), German (Chmitorz *et al*., 2018), Greek (Kyriazos *et al*., 2018), Chinese (Lai & Yue, 2014) and Japanese (Iimura & Taku, 2018). As stated, The Brief Resilience Scale, which is used in many cultures, is widely used in research, but there are problems with its factor structure. Studies support both one-factor and two-factor models (Crane & Searle, 2016; Tansey *et al*., 2016). For this reason, testing the factor structure of the scale was deemed important and it was aimed to examine it within the scope of the Rasch model.

METHOD

The objective of this study is to assess the psychometric properties of the Brief Resilience Scale within the context of the Rasch Partial Credit model. Accordingly, the study group, data collection tool, and data analysis are given as follows.

Study Group

The minimum sample size proposed for the studies conducted within the scope of the Rasch model was determined to be at least 500. In cases where the sample size is small, parameter estimates will be inaccurate, and it will be challenging to ensure the invariance of item-ability parameters (de Ayala, 2009; DeMars, 2010; Zanon *et al*., 2016). A total of 608 individuals were approached by paying attention to the minimum value specified within the scope of this research study. The study group of the research consists of 608 students studying at a public university in Istanbul. 63.9% of the individuals were female (N=389), and 36.1% were male (N=219). 69.4% of the study group were between the ages of 18 and 25. 30.6% were individuals over the age of 25. The age-related mode was 22, with an average of 26.11. The privacy of all the participants was protected, and confidentiality requirements for data collection and analysis were strictly followed.

Data Collection Tool

In this study, the Brief Resilience Scale, which is widely used in social sciences, was preferred.

Brief Resilience Scale (BPRS): The scale developed by Smith *et al*. (2008) was created to measure the resilience of individuals. The BPRS is a five-point Likert-type measurement tool consisting of six items. High scores obtained from the scale indicate high resilience. The development of the scale and validity and reliability studies were carried out on four different study groups. Accordingly, the first two groups consisted of university students, and the next two groups consisted of cardiac and fibromyalgia patients. Exploratory factor analysis was performed to determine the construct validity of the scale; and as a result of the analysis, a single-factor structure explaining 61%, 61%, 57%, and 67% of the total variance was obtained for four different sample groups, respectively. The factor loadings of the scale items ranged between .68 and .91. The reliability of the scale was calculated by internal consistency and test-retest methods; the internal consistency reliability coefficient was found to be between .80 and .91. The test-retest reliability coefficient was found between .62 and .69. Within the scope of criterion-related validity, the relationships between the BPRS and other scales were examined. Accordingly, significant positive correlations were found between BPRS and ego resilience, optimism, life goals, social support, positive coping strategies and positive emotions. On the other hand, significant negative correlations were found between the BPRS and pessimism, depression, anxiety, negative emotions, perceived stress and negative coping strategies.

In the adaptation conducted by Doğan (2015), 295 (186 female, 109 male) university students were studied. As a result of the exploratory factor analysis, a one-factor structure explaining 54% of the total variance was obtained, and factor loadings ranged from .63 to .79. As a result of DFA conducted in parallel with BPRS, goodness-of-fit indices were found as x^2/sd (12.86/7) = 1.83, NFI = 0.99, NNFI = 0.99, CFI = 0.99, IFI = 0.99, RFI = 0.97, GFI = 0.99, AGFI = 0.96, RMSEA = 0.05, SRMR = 0.03. The reliability of the BPRS was examined with the internal consistency method and accordingly, the internal consistency coefficient was found to be .83. The BPRS is a five-point Likert scale and has response categories such as "not appropriate at all" (1), "not appropriate" (2), "somewhat appropriate" (3), "appropriate" (4), and "completely appropriate" (5). Items 2, 4, and 6 are reverse items.

Data Analysis

The data gathered for the research study were analyzed in the jamovi2.3.13 program. Unidimensionality, local independence, and model-data fit are the assumptions necessary to carry out the Rasch analysis (DeMars, 2010). Unidimensionality is the presence of a single latent property that adequately explains the common variance, and the observed variables are a function of only a single latent variable (de Ayala, 2009; Embretson & Reise, 2000). In this study, the unidimensionality assumption was tested using Principal-Axis Factor Analysis. A principal-axis factor analysis was implemented since it is considered one of the best methods by which the nature of the underlying structure behind the items can be understood (Costello & Osborne, 2005). Meeting the unidimensionality assumption also demonstrates that there is no problem with local independence (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985; Hambleton *et al*., 1991; Lord, 1980; Morizot *et al.*, 2007). Local independence means that the items are independent of one another for a certain skill level. Although it is stated that if the unidimensionality assumption is provided, the local independence assumption will also be provided, it is recommended reviewing the local independence assumption as well (DeMars, 2010). Local independence of items is an assumption in Rasch model, meaning the items in a test should not be related to each other (Baghaei, 2008). A violation of the assumption of local independence may occur if the response to one item affects another item and the scale is multidimensional. Standardized residual correlations reflect local dependency between items, where correlations above .7 suggest that two items share more than half of their random variance and only one needs to be retained (Linacre, 2019). Accordingly, Yen's (1984) Q3 statistic was utilized to test the local independence assumption. Expressed as a correlation coefficient of residual values between items, the Q3 statistic is a statistic that indicates the dependence between item pairs. To evaluate the degree of local independence, the relationship between all possible pairs of items needs to be examined. While .20 criteria are used in the evaluation of Yen's Q3 statistic, (Christiensen *et al*., 2017), .30 criteria (Røe *et al*., 2014) are taken into account as well.

Item fit statistics were assessed with infit and outfit values. The infit value is more sensitive to the individual's responses to items with similar difficulty levels and provides central information. The outfit value, on the other hand, is more sensitive to the unexpected responses of the individual to the more difficult or simpleritems. Infit and Outfit take values ranging from 0 to ∞, yet the value indicating perfect fit is 1.00 (Eckes, 2009). Nevertheless, it is hard to find a perfect fit between the model and the data (Brentari & Golia, 2008). These two values are evaluated together, and a value between 0.50 and 1.50 indicates that item-model fit is achieved (Linacre, 2019). Another advantage of the Rasch model is that it is possible to determine how well the scale categories work (Linacre, 2014). Thus, the category statistics in the Rasch analysis outputs and whether the response categories for each item were in order were investigated. The item characteristic curve of the scale items was obtained as well. Item characteristic curves show expected and observed characteristic curves, and, observed and expected characteristic curves are expected to correspond for model fit (Yan & Heene, 2020).

Person Reliability was utilized for reliability within the context of the Rasch model. It is known that the closer Person Reliability is to 1.00, as in other reliability coefficients, the higher the reliability. This coefficient is also known as Person Separation Reliability (Chang *et al*. 2014; de Ayala, 2009). It is suggested that .70 be taken as a criterion for the reliability index achieved from the Rasch analysis, as in the Cronbach alpha internal consistency coefficient (Walker *et al*., 2012).

Based on the data obtained, it was investigated whether the items showed Differential Item Functioning (DIF). DIF measuring means the matching of individuals with respect to their abilities in terms of the variable to be measured and afterwards statistically determining that these individuals in different groups have different probabilities of responding to the item (Camilli & Shepard, 1994; Clauser & Mazor, 1998; Roever, 2005; Zumbo, 1999). There are two types of DIF; uniform and non-uniform. If Uniform DIF is present, the difference between the item characteristic curves for the focus and reference groups is uniform (Finch & French, 2007; Jodoin & Gierl, 2001; Walker, 2011). Non-uniform DIF occurs when the difference between the substance characteristic curves is not constant (Walker *et al*., 2001). As a result of statistical analysis, items are labelled in the A (nonsignificant DIF), B (moderate DIF), and C (high DIF) categories (Zieky, 1993). In this research study, the Likelihood Ratio method was utilized to establish DIF according to gender within the scope of the Rasch model and multi-category scoring. In this method, the hypothesis of whether there is a difference between the focus and reference group item parameters is tested; also, limited and generalized models are created, and and their ratios to one another are tested by creatin accordingly (Atalay *et al*., 2012). By taking the Likelihood Ratio logarithm, the G^2 value is obtained and checked from the Chi-Square table using the degrees of freedom. If this value is significant, this shows the presence of DIF (Thissen, 2001). G^2 values give information about DIF size. The DIF levels are presented below based on the values of the G^2 value (Greer, 2004; Thissen, 2001):

If it is $3.84 < G^2 < 9.4$, then, no DIF or DIF at a negligible level,

$9.4 \leq G^2 < 41.9$ shows a middle level of DIF,

$G^2 \geq 41.9$ shows a high level of DIF.

FINDINGS

For the Partial Credit Rasch Model, the assumptions were tested first. Principal-Axis Factor Analysis was applied to test the unidimensionality assumption. The KMO value was .86, and the Barlett's Test of Sphericity result was significant ($\chi2$= 1713, *df*= 15). This showed that the data was suitable for factor analysis. As a result of the Principal-Axis Factor Analysis, it was determined that there was a single factor structure (Figure 1). Factor loads are values of between .74 and .83. The explained variance of the single factor structure was determined to be 61.20%. According to Linacre (2019), a scale would be considered unidimensional if the primary dimension explains at least 40% of the variance.

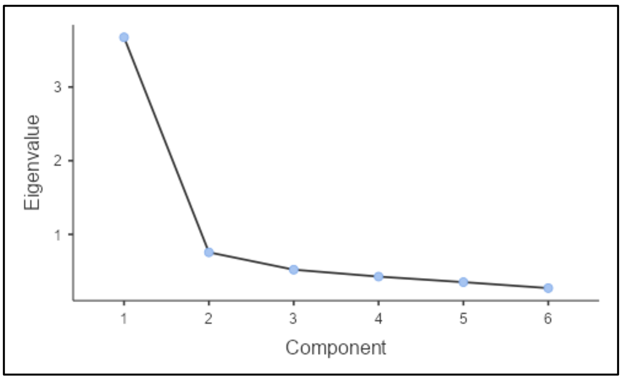


Figure 1. Scree plot for scale

Although it was stated that if the unidimensionality assumption is met, the local independence assumption will also be met, The Yen's Q3 statistics were used for the local independence assumption. Thus, following the parameter estimation, a residual matrix was created using the residue of each item, and correlations between them were examined. The Q3 Correlation Matrix is presented in Table 1.

Table 1. Q3 Correlation Matrix

	Item1	Item2	Item3	Item4	Item5	Item6
Item 1	—					
Item 2	-0.241	—				
Item 3	-0.042	-0.116	—			
Item 4	-0.280	-0.126	-0.296	—		
Item 5	-0.083	-0.276	-0.127	-0.245	—	
Item 6	-0.202	-0.167	-0.294	0.125	-0.203	—

Examining the Q3 correlation matrix, it is seen that the Q3 correlations between the item pairs (as an absolute value) are less than .30. Researchers stated that item pairs with values above .30 violated local independence. Considering this criterion, it is seen that the assumption of local independence is achieved. As noted by Baghaei (2008) local independence of items is an assumption in Rasch model, meaning the items in a test should not be related to each other. Therefore, it was revealed that the responses to the six items in the scale were not related to each other. After the assumptions of the Rasch model were met, a model comparison was made to determine the model that was suitable for the data. The results of the Partial Credit Model (PCM) and Rating Scale Model (RSM) comparison are given in Table 2.

Table 2. Model Comparision

Model	**Log-likelihood**	**AIC**	**BIC**	**CAIC**	**Parameters**	**N**	**χ^2**	**df**	**p**
PCM	-4654	9358	9423	9493	25	608	50.5	15	< .001
RSM	-4679	9379	9468	9433	10	608			

Table 2 shows the Log-likelihood, AIC, BIC and CAIC values used for model comparison. Janssen & De Boeck (1999) stated that the model with low log-likelihood and fit values is more compatible with the data. Therefore, χ^2= 50.5, *df*= 15; *p*<.001 indicates that there is a difference between the two models and PCM is more compatible with the data. Obtained by PCM, Infit and Outfit values for item-model fit are shown in Table 3.

Table 3. Item statistics of the model

	Item Mean	Measure	SE Measure	Infit	Outfit	Point Biserial
Item 1	2.43	-0.659	0.0612	0.953	0.922	0.774
Item 2	2.06	0.161	0.0567	0.967	1.009	0.773
Item 3	2.32	-0.512	0.0572	1.144	1.141	0.731
Item 4	2.02	0.264	0.0578	0.940	0.931	0.783
Item 5	1.82	0.551	0.0600	1.223	1.211	0.726
Item 6	2.05	0.195	0.0562	0.822	0.787	0.807

Infit = Information-weighted mean square statistic
Outfit = Outlier-sensitive means square statistic

The item difficulty parameters present the position of items on the latent trait continuum and are explained in logits (or log odd-units). The error of measurement shows to what extent the item difficulty parameters were accurately estimated. Infit and outfit values give item fit statistics and are expected to be in the range of .50 to 1.50. The obtained fit values vary between .82 - 1.22, and the Outfit values vary between .78 - 1.21. Eckes (2009) states that the desired value is 1.00, but it is very difficult to achieve this perfect fit. Wright & Linacre (1994) interpreted values between .50 and 1.50 for fit statistics as acceptable fit criteria. Therefore, as seen in Table 3, six items appear to be compatible with the model.

Linacre (2009) stated that 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. Point-biserial (or point-measure) correlations indicate to what extent the responses to each item within a measure are correlated with the overall measure. Therefore, it is stated that a value of .30 can be taken as a criterion for this correlation coefficient, as in item discrimination (Güler *et al*., 2017). As can be seen from Table 3, the point biserial values of the items vary between .72 and 0.80. Another advantage of the Rasch model is that it allows determining how well the response categories work (Linacre, 2014). The threshold values for the response categories of the items are presented in Table 4.

Table 4. Thresholds of the partial credit model

	Threshold 1	Threshold 2	Threshold 3	Threshold 4
Item 1	-1.614	-1.1187	0.851	2.60
Item 2	-0.981	0.0115	1.114	3.31
Item 3	-1.971	-0.1206	0.767	2.49
Item 4	-0.961	-0.0702	1.251	3.58
Item 5	-1.215	0.3520	1.958	3.59
Item 6	-0.711	-0.1470	1.211	3.22

Because an item has K ordered option responses, PCM estimates K − 1 thresholds for the item. PCM does not require the thresholds to follow the same order as the response categories. In Table 4, it can be seen that threshold parameters also increase in response categories from the lowest to the highest, indicating that the five-point grading in the scale works effectively. As Linacre (2002) stated; sequential increase indicates that people with more latent variables are more likely to choose categories higher on the scale. Item characteristic curves for ten items in the scale are given in Figure 2 and the item category for PCM in Figure 3.

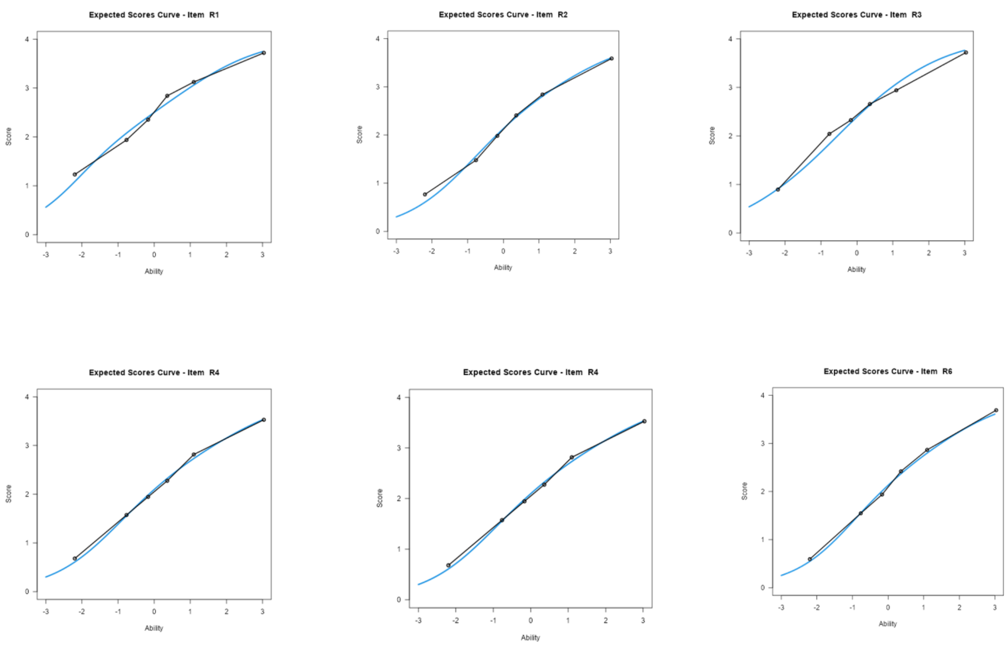


Figure 2. *Item characteristic curve*

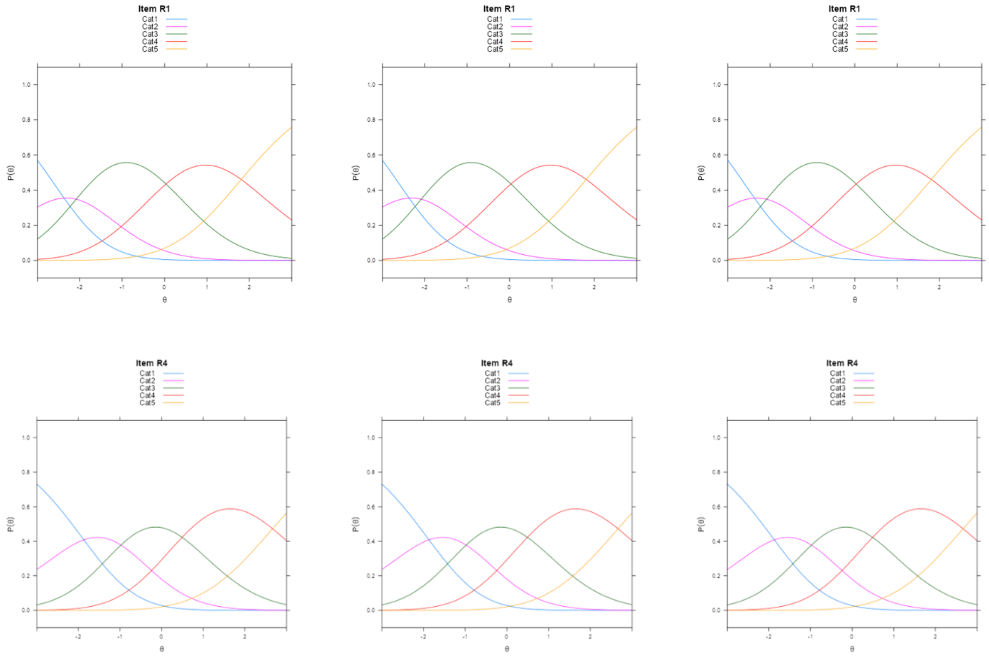


Figure 3. *Item category for PCM*

For the Item Characteristic Curve, the blue line is a theoretical item characteristic curve obtained using parameter b. The black line, on the other hand, shows the responses to this item of the individuals for whom talent estimation was made with respect to the data included in the analysis. Reviewing the item characteristic curve for the items, it is seen that the actual probabilities are compatible with the expected probability. Examining the Item category for PCM, it can be that there is a sequential transition from the "not appropriate at all" response to the "completely appropriate" response to the item on the horizontal axis as the skill level increases. The Wright Map is also presented in Figure 4.

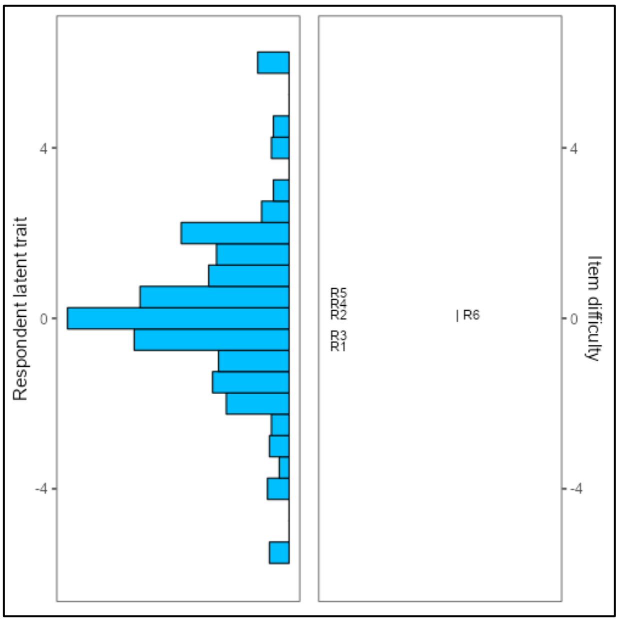


Figure 4. *Wright Map*

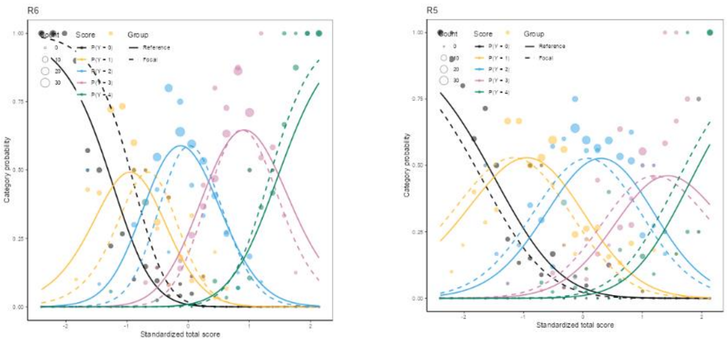
Person-item maps, also named Wright Maps, show the distribution of item difficulties and the distribution of individuals' responses (Linacre, 2008). On the left side of the graph, the histogram graph of the individuals' ability estimations can be seen, and on the right side, the distribution of the items according to their difficulties can be seen. The upper part of the graph shows high ability levels and the lower part shows low ability levels. Based on the graph, it is seen that R2 and R6 have similar difficulty levels.

The Person Reliability value obtained within the scope of the Rasch model was found to be .86, and this value shows that the reliability is high. It is recommended that .70 be taken as a criterion for this value, which is interpreted similarly to the Cronbach alpha internal consistency coefficient (Bond & Fox, 2012; Walker *et al*., 2012). Therefore, in line with the obtained value, it has been demonstrated that the result obtained according to the Rasch model is reliable.

DIF was reviewed with respect to the outcomes obtained from the six-item scale. In this scope, the Likelihood Ratio method, one of the DIF determination methods, was utilized, and the results are given in Table 5 and Figure 5.

Table 5. *DIF results according to the Likelihood Ratio Method*

	Non-uniform		Uniform	
	G^2	*p*	G^2	*p*
Item 1	0.0696	0.792	0.0122	0.912
Item 2	2.3282	0.127	0.7580	0.384
Item 3	1.3794	0.240	1.1374	0.286
Item 4	2.4411	0.118	1.2591	0.262
Item 5	0.0837	0.772	4.5295	0.033
Item 6	4.2066	0.040	2.5383	0.111



Non-uniform - Item 6 Uniform – Item 5

Figure 6. *Item characteristic curve for reference and focal group*

DIF occurs when an items properties in one group are different from the item's properties in another group (Furr & Bacharach, 2017). Examining Table 5, the sixth item shows DIF within the scope of Non-uniform DIF, and the G^2 value is calculated to be 4.21. Based on the likelihood ratio statistical values, a G^2 value in the range of

$3.84 < G^2 < 9.4$ indicates a negligible (A) DIF (Greer, 2004). Therefore, it can be said that the sixth item has a negligible level of DIF. Within the context of Uniform DIF, the fifth item shows DIF, and the G^2 value is calculated as 4.53. It can be said that this item also shows a negligible level of DIF. When Figure 5 is examined, it can be seen that the sixth and fifth items function differently for the reference and focal groups.

DISCUSSION and SUGGESTIONS

In this study, the psychometric properties of the Brief Resilience Scale, which was developed by Smith *et al*. (2008), adapted by Doğan (2015), and widely used in both education and psychology, were tested within the scope of Rasch PCM. Studies comparing IRT and CTT state that IRT has stronger properties (Lord, 1980; Blood, 2006; Gelbal, 1994; Doğan & Tezbaşaran, 2003; Nartgün, 2002). Therefore, it was found noteworthy to test the psychometric properties of the Brief Resilience Scale within the scope of Rasch PCM and interpret the results. Principal-Axis Factor Analysis was carried out to test the unidimensionality in line with the assumptions of the Partial Credit Rasch Model. Researchers (Brown, 2015) emphasize that this method is generally utilized to control unidimensionality. With respect to the results, a one-dimensional structure with an explained variance of 61.20% was obtained. It is stated that the variance explained in the social sciences should be at least 40% and above (Stevens, 1992). The developers of the scale also found that a single-factor structure had an acceptable fit using Principal Components Analysis with four small samples ranging between 50 and 128 participants (Smith *et al*., 2008). Doğan (2015) also reported the explained variance of the unidimensional resilience structure as 54%. Liu & Lim (2020), who studied the scale within the scope of the Rasch model, also obtained an explained variance close to 50%. In studies conducted within the scope of CTT, Coelho *et al.* (2016) confirmed the unidimensional structure with 43% explained variance. Unlike these studies, Crane & Searle (2016), Kyriazos *et al*. (2018), Tansey *et al*. (2016) and Fung (2020) stated in their studies that the two-factor structure is more suitable for the Brief Resilience Scale. Finally, McKay, Skues & Williams (2021) the researchers asked “does the Brief Resilience Scale actually measure resilience and succumbing?” and they reported in their study titled that the unidimensional scale had better fit indices.

For the local independence assumption, Yen's Q3 statistics were utilized, and it was established that the Q3 correlations between the item pairs were less than .30, and therefore a response to one item did not affect another item. Marais (2009) and Yen (1993) suggested that if the local independence assumption cannot be met, parameter estimations based on individuals may affect the reliability and validity of the scale results. Therefore, the fact that local independence was achieved with the data obtained from this study also contributed to more accurate estimations.

It was revealed that the Infit values testing the item fit statistics ranged from .82 to 1.22, and the Outfit values ranged from .78 to 1.21, and that all the items were compatible with the model. The outfit value is much more sensitive to unexpected extremes than the infit statistic. Researchers (Güler, 2014; Hetherman, 2004) emphasize that an infit value above 1.00 indicates a higher-than-expected variance, and a value lower than 1.00 indicates a lower-than-expected variance. The fact that the fit statistic average is quite close to 1.00 indicates that the model-data consistency is quite high. Based on both the findings of this research study and the findings of Liu & Lim (2020), working within the scope of the Rasch model, it was revealed that all the items were compatible with the model. It was observed that threshold parameters elevated in response categories from the lowest to the highest, indicating that the five-point grading in the scale works effectively. Liu & Lim (2020) also revealed that the five-point grading system in the scale is effective. All categories were sufficiently used to provide precise and stable estimates, and the noise observed did not degrade the measurement.

The point biserial values of the items were found to vary within the range of .72 and .80. According to Güler *et al*. (2017) and Linacre (2014), since the values are greater than .30, it has been established that all items function in the same direction and should remain on the scale. When the item characteristic curve for the items was examined, it was established that the actual probabilities were compatible with the expected probability. Walker, Engelhard, and Thompson (2012) suggested that .70 should be taken as a criterion for the reliability index obtained from the Rasch model, as in the Cronbach alpha internal consistency coefficient. The Person Reliability value obtained within the scope of the Rasch model was found to be .86, and this value shows that the reliability is high. Doğan (2015) also reported the internal consistency coefficient of the scale as .83. Liu & Lim (2020) also reported the Person Reliability value to be 0.77. In studies conducted within the scope of CTT, Rodríguez-Rey *et al.* (2015) found validity and reliability for the Brief Resilience Scale at the .83 level, while Chmitorz *et al*. (2018) achieved reliability at the .85 level, Coelho *et al.* (2016) at the .77 level, Kyriazos *et al*., (2018) at the .80 level, and Fung (2020) at the .71 level.

Based on the results obtained from the six-item scale, DIF was investigated in terms of gender; and according to the likelihood ratio statistical values, it was determined that the fifth and sixth items showed a negligible level of DIF. Liu & Lim (2020) similarly determined a negligible level of DIF by gender in items 5 and 6. In a study where the psychometric properties of the Resilience Scale (RS-25) were determined, it was examined whether the items showed DIF according to gender and evidence was obtained that there was no DIF according to gender (Seong *et al*., 2023). In their study where Chen and colleagues (2020) examined DIF according to gender with the Chinese version of the Resilience Scale (RS-14), they detected DIF according to gender in four items of the scale.

Despite compelling evidence for Rasch analysis as a robust scale validation method, there is only one study that has evaluated the BRS based on Rasch. Therefore, studying the data obtained from Turkey within the scope of the Rasch model has been informative for both the literature and practitioners. In line with the explanations above, the findings of this study should be assessed within the framework of some limitations. The majority of the individuals in the study group were female. DIF was only determined by the likelihood ratio method. A second method was not utilized as a negligible level of DIF was determined; however, researchers (Karami & Nodoushan, 2011) state that using a single technique and basing the results on a single technique may not be appropriate. Therefore, a second DIF determination method can be used, and expert opinions can be obtained to interpret item bias. In this research study, the scale was evaluated within the scope of the Rasch model, but it could be evaluated within the scope of Generalizability Theory in other studies as well. Despite all these limitations, it has been demonstrated that the Brief Resilience Scale provides valid and reliable results within the scope of Rasch PCM and is unidimensional.

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. Ethics committee permission for the article was received by Istanbul Okan University / Publication Ethics Board with the decision numbered 134 dated 10.03.2021.

Author Contribution Rate: The author's contribution rate in this research is 100%.

REFERENCES

Abel, M. H., & Sewell, J. (1999). Stress and burnout in rural and urban secondary school teachers. *The Journal of Educational Research*, *92*(5), 287-293.

Adedoyin, O. O. (2010). Investigating the invariance of person parameter estimates based on classical test and item response theories. *International Journal of Educational Sciences*, *2*(2), 107-113.

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. Retrieved from https://dergipark.org.tr/en/pub/auad/issue/34247/378491.

Alavi, K., Isa, K., & Palpanadan, S. T. (2020). Application of Rasch model on resilience in higher education: an examination of validity and reliability of Malaysian Academician Happiness Index (MAHI). *International Journal of Higher Education*, *9*(4), 261-271.

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

Ardelt, M., Ai, A. L., & Eichenberger, S. E. (2008). In search for meaning: The differential role of religion for middle-aged and older persons diagnosed with a life-threatening illness. *Journal of Religion, Spirituality & Aging*, *20*(4), 288–312. https://doi.org/10.1080/15528030802232353

Atalay, K., Gök, B., Kelecioğlu, H. & Arsan, N. (2012). Değişen madde fonksiyonunun belirlenmesinde kullanılan farklı yöntemlerin karşılaştırılması: Bir simülasyon çalışması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 43*, 270-281.

Baghaei, P. (2008). Local dependency and Rasch measures. *Rasch Measurement Transactions*, *21*(3), 1105–1106. Retrieved from https://www.rasch.org/rmt/rmt213b.htm.

Benard, B. (1991). *Fostering resiliency in kids: Protective factors in the family, school and community*. Northwest Regional Educational Laboratory.

Benard, B. (2004). *Resiliency: What we have learned*. WestEd.

Bernshausen, D., & Cunningham, C. (2001). The role of resiliency in teacher preparation and retention. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Dalas, TX (ERIC Document Reproduction Service No. ED 451 191).

Bond, T.G., & Fox, C.M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Erlbaum.

Boone, WJ. (2016). Rasch analysis for instrument development: why, when, and how? *CBE-Life Sciences Education*, *15*(4):1-7.

Brentari, E., & Golia, S. (2008). Measuring job satisfaction in the social services sector with the Rasch model. *Journal of Applied Measurement*, *9*(1), 45-56. Retrieved from http://www.unibs.it/sites/default/files/ricerca/allegati/10061.pdf

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, S., & Nagel, L. (2004). Preparing future teachers to respond to stress: Sources and solutions. *Action in Teacher Education*, *26*(1), 34-42.

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.

Camilli G., & Shepard L. A. (1994). *Methods for identifying biased test items* (volume 4). SAGE Publications.

Çelen, Ü. (2008). Klasik test kuramı ve madde tepki kuramı yöntemleriyle geliştirilen iki testin geçerlilik ve güvenilirliğinin karşılaştırılması. *İlköğretim Online*, *7*(3), 758-768.

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.

Chmitorz, A., Wenzel, M., Stieglitz, R. D., Kunzler, A., Bagusat, C., Helmreich, I., ... & Tüscher, O. (2018). Population-based validation of a German version of the Brief Resilience Scale. *PloS one*, *13*(2), e0192761.

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.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differential item functioning test items. *Educational Measurement*: Issues and Practice, *17*, 31-44.

Coelho, G.L.D.H., Hanel, P.H.P., Cavalcanti, T.M., Rezende, A.T., & Gouveia, V.V. (2016). Brief resilience scale: Testing its factorial structure and invariance in Brazil. *Universitas Psychologica*, *15*(2), 379–408.

Connor, K.M., & Davidson, J.R.T. (2003). Development of a new resilience scale: The Connor-Davidson resilience scale (CD-RISC). *Depression and Anxiety*, *18*(2), 76–82.

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.

Coutu, D. L. (2002). How resilience works. *Harvard business review*, *80*(5), 46-56.

Crane, M. F., & Searle, B. J. (2016). Building resilience through exposure to stressors: The effects of challenges versus hindrances. *Journal of Occupational Health Psychology*, *21*(4), 468–479. doi:10. 1037/a0040064

Dapueto, J. J., Francolino, C., Servente, L., Chang, C. H., Gotta, I., Levin, R., & Abreu, M. D. C. (2003). Evaluation of the Functional Assessment of Cancer Therapy-General (FACT-G) Spanish Version 4 in South America: classic psychometric and item response theory analyses. *Health and quality of life outcomes*, *1*(1), 1-12.

Day, C., & Gu, Q. (2014). *Resilient teachers, resilient schools: Building and sustaining quality in testing times*. Routledge.

Day, C., Edwards, A., Griffiths, A., & Gu, C. (2011). *Beyond survival teachers and resilience*. Key Messages from an ESRC-funded Seminar Series. The Univeristy of Nottingham.

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

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

Doğan, T. (2015). Kısa psikolojik sağlamlık ölçeği'nin Türkçe uyarlaması: Geçerlik ve güvenirlik çalışması. *The Journal of Happiness & Well-Being*, *3*(1), 93-102.

Earvolino-Ramirez, M. (2007). Resilience: A concept analysis. In *Nursing forum* (Vol. 42, No. 2, pp. 73-82). Blackwell Publishing Inc.

Eckes, T. (2009). Many-facet Rasch measurement. In S. Takala (Ed.), *Reference supplement to the manual for relating language examinations to the Common European Framework of Reference for Languages: Learning, teaching, assessment (Section H).* Council of Europe/Language Policy Division.

Ee, J., & Chang, A. S. C. (2010). How resilient are our graduate trainee teachers in Singapore? *The Asia-Pacific Education Researcher*, *19*(2), 321-331.

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.

Faigin, C. A., & Pargament, K. L. (2011). Strenghtened by the spirit: Religion, spirituality, and resilience through adulthood and aging. In B. Resnick, L. P. Gwyther, & K. A. Roberto (Eds.), *Resilience in aging: Concepts, research and outcomes* (pp. 163–178). Springer.

Fimian, M. J. (1984). The development of an instrument to measure occupational stress in teachers: The teacher stress inventory*. Journal of Occupational Psychology*, *57*(4), 277-293.

Finch, W. H., & French F. B. (2007). Detection of crossing differential item functioning: a comprasion of four methods. *Educational and Psychological Measurement*, *67*(4), 565-582.

Friborg, O., Hjemdal, O., Rosenvinge, J.H., & Martinussen, M. (2003). A new rating scale for adult resilience: What are the central protective resources behind healthy adjustment? *International Journal of Methods in Psychiatric Research*, *12* (2), 65–76.

Fung, S. F. (2020). Validity of the brief resilience scale and brief resilient coping scale in a Chinese sample. *International journal of environmental research and public health*, *17*(4), 1265.

Furr, M.R., & Bacharach, V.R. (2017). *Psychometrics: An Introduction* (2nd ed.). SAGE Publications, Inc.

Garrett, P. (2009). *A Monte Carlo study ınvestigating missing data, differential item functioning and effect size* (Doctoral Dissertation). Georgia State University.

Giroux, P. (2007). *Resilient teachers: A qualitative study of six thriving educators in urban elementary schools.* Western Michigan University, Michingan.

Goolmohammadnejhad, G. (2009). Introduction and analytical comparison of classic test theory (CTT) and item response theory (IRT) in psychometrics. *The Journal of Productivity Management*, *2*(4), 217-244.

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. (2014). Analysis of Open-Ended Statistics Questions with Many Facet Rasch Model. *Eurasian Journal of Educational Research*, *55*, 73-90.

Güler, N., & Gelbal, S. (2010). A study based on classic test theory and many facet Rasch model. *Eurasian Journal of Educational Research*, *38*, 108-125.

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

Guo, Y., Liu, S., Zhou, F., Hu, Z., Li, S., & Xu, H. (2023). Psychometric properties of the Chinese version of the Nursing Home Adjustment Scale (C-NHAS) in Central China: Classical test theory and item response theory. *Geriatric Nursing*, *51*, 102-111.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. CA: Sage.

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

Henderson, N., & Milstein, M. M. (2003). *Resiliency in schools: Making it happen for students and educators*. Corwin press.

Hetherman, S. C. (2004). *An application of multi-faceted Rasch measurement to monitor effectiveness of the written composition in English in the New York City Department of Education*. Teachers College, Columbia University.

Iimura, S., & Taku, K. (2018). Gender differences in relationship between resilience and big five personality traits in Japanese adolescents. *Psychological Reports*, *121*(5), 920–931.

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.

Janssen, R., & De Boeck, P. (1999). Confirmatory analyses of componential test structure using multidimensional item response theory. Multivariate Behavioral Research, *34*(2), 245–268. doi:10 . 1207 / S15327906Mb340205

Jefferies, P., McGarrigle, L., & Ungar, M. (2019). The CYRM-R: A Rasch-validated revision of the child and youth resilience measure. *Journal of Evidence-Based Social Work*, *16*(1), 70-92.

Jodoin, M. G., & Gierl, M. J. (2001). Evaluating Type I error and power rates using an effect size measure with logistic regression procedure for DIF detection. *Applied Measurement in Education*, *14*, 329–349.

Karami, H., & Salmani Nodoushan, M. A. (2011). Differential Item Functioning (DIF): Current Problems and Future Directions. *Online Submission*, *5*(3), 133-142.

Keshavarz Gerami, G., Akbari Balootbangan, A., & Babakhani, K. (2017). Comparison of the psychometric characteristics of the adolescents depression scale based on the ıtem response theory with those based on the classic test theory. *Journal of School of Public Health and Institute of Public Health Research*, *15*(3), 209-224.

Kim, H. B., Nam, J. H., Chung, D. H., & Cho, E. H. (2022). Developing a scale using item response theory of the self-resilience in taekwondo players. *International journal of environmental research and public health*, *20*(1), 728.

Kobasa, S. C. (1979). Stressful life events, personality, and health: An inquiry into hardiness. *Journal of Personality and Social Psychology*, *37*(1), 1-11.

Kobasa, S. C., & Puccetti, M. C. (1983). Personality and social resources in stress resistance. *Journal of Personality and Social Psychology*, *45*(4), 839.

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.

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.

Kyriazos, T.A., Stalikas, A., Prassa, K., Galanakis, M., Yotsidi, V., & Lakioti, A. (2018). Psychometric evidence of the Brief Resilience Scale (BRS) and modeling distinctiveness of resilience from depression and stress. *Psychology*, *9*(7), 1828–1857.

Lai, J.C.L., & Yue, X. (2014). *Using the brief resilience scale to assess Chinese people's ability to bounce back from stress*. SAGE Open.

Lee, M. K., Peipert, J. D., Cella, D., Yost, K. J., Eton, D. T., Novotny, P. J., ... & Dueck, A. C. (2023). Identifying meaningful change on PROMIS short forms in cancer patients: a comparison of item response theory and classic test theory frameworks. *Quality of Life Research*, *32*(5), 1355-1367.

Li, C., Li, S., & Wei, C. (2019). The development and validation of Academic Resilience Scale for undergraduates in Taiwan: Rasch analysis. In *2nd International Conference on Humanities Education and Social Sciences (ICHESS 2019)* (pp. 205-208). Atlantis Press.

Lin, C. Y., Imani, V., Griffiths, M. D., & Pakpour, A. H. (2021). Validity of the Yale Food Addiction Scale for Children (YFAS-C): Classical test theory and item response theory of the Persian YFAS-C. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, *26*, 1457-1466.

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. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*(1), 85-106. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.424.2811&rep=rep1&type=pdf

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. (2019). *A user's guide to Winsteps® Rasch-model computer programs*. Beaverton, Oregon.

Liu, V. Y. Y., & Lim, S. M. (2022). A psychometric evaluation of the brief resilience scale among tertiary students in Singapore. *Asia Pacific Journal of Education*, *42*(3), 464-477

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

Maddi, S. R., & Khoshaba, D. M. (2005). *Resilience at work: How to succeed no matter what life throws at you*. Amacom.

Manning, L., Ferris, M., Rosario, C. N., Prues, M., & Bouchard, L. (2019). Spiritual resilience: Understanding the protection and promotion of well-being in the later life. *Journal of Religion, Spirituality & Aging, 31*(2), 168-186. https://doi.org/10.1080/15528030.2018.1532859

Marais, I. (2009). Response dependence and the measurement of change. *Journal of Applied Measurement*, *10*(1), 17-29.

Masten, A. S., & Reed, M.-G. (2002). Resilience in development. In C. R. Snyder & S. J. Lopez (Eds.), Handbook of positive psychology (pp. 74-88). Oxford University Press.

Masten, A., Best, K., & Garmezy, N. (1990). Resilience and development: Contributions from the study of children who overcome adversity. *Developments in Psychopathology, 2*, 425–444.

McKay, S., Skues, J. L., & Williams, B. J. (2021). Does the Brief Resilience Scale actually measure resilience and succumbing? Comparing artefactual and substantive models. *Advances in Mental Health*, *19*(2), 192-201.

Morizot, J., Ainsworth, A. T., & 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.

Nartgün, Z. (2002). Aynı tutumu ölçmeye yönelik Likert tipi ölçek ile metrik ölçeğin madde ve ölçek özelliklerinin klasik test kuramı ve örtük özellikler kuramına göre incelenmesi. Yayımlanmamış doktora tezi [The investigation of item and scale properties of Likert type scale and metric scale measuring the same attitude according to classisical test theory and item response theory], Hacettepe Üniversitesi.

Neenan, M. (2009). *Developing resilience: A cognitive-behavioural approach*. Routledge.

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.

Peres, J. F., Mercante, J., & Nasello, A. G. (2005). Psychological dynamics affecting traumatic memories: implications in psychotherapy. *Psychology and psychotherapy: theory, research and practice, 78*(4), 431-447. https://doi.org/10.1348/147608305X26693

Peres, J. F., Mercante, J., & Nasello, A. G. (2005). Psychological dynamics affecting traumatic memories: implications in psychotherapy. *Psychology and psychotherapy: theory, research and practice, 78*(4), 431-447. https://doi.org/10.1348/147608305X26693

Peres, J. F., Moreira-Almeida, A., Nasello, A. G., & Koenig, H. G. (2007). Spirituality and resilience in trauma victims. *Journal of religion and health*, *46*, 343-350. https://doi.org/10.1007/sl0943-006-9103-0

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

Ricketts S. N., Engelhard G., & Chang M. L. (2015) Development and validation of a scale to measure academic resilience in mathematics. *European Journal of Psychological Assessment*, *33*(2): 79–96. doi:10.1027/1015-5759/a000274.

Rodríguez-Rey, R., Alonso-Tapia, J., & Hernansaiz-Garrido, H. (2016). Reliability and validity of the brief resilience scale (BRS) Spanish version. *Psychological assessment*, *28*(5), e101.

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, PubMed 24646065

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

Sait, Ç., & 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.

Schuster, M. A., Stein, B. D., Jaycox, L. H., Collins, R. L., Marshall, G. N., Elliott, M. N., ... & Berry, S. H. (2001). A national survey of stress reactions after the September 11, 2001, terrorist attacks. *New England Journal of Medicine*, *345*(20), 1507-1512.

Simmons, K. H. (2011). *Overcoming the odds: An exploration of resilience strategies used by female superintendents in school districts in Georgia* (Doctoral dissertation). Georgia Southern University, Statesboro.

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: assessing the ability to bounce back. *International journal of behavioral medicine*, *15*, 194-200.

Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Second Edition, Lawrence Erlbaum Associates.

Tansey, T. N., Kaya, C., Moser, E., Eagle, D., Dutta, A., & Chan, F. (2016). Psychometric validation of the Brief Resilience Scale in a sample of vocational rehabilitation consumers. *Rehabilitation Counseling Bulletin, 59*(2), 108–111. doi:10.1177/0034355215573539

Teker, G. T., Guler, N., & Uyanik, G. K. (2015). Comparing the Effectiveness of SPSS and EduG Using Different Designs for Generalizability Theory. *Educational Sciences: Theory and Practice*, *15*(3), 635-645.

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.

Walker C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and validation. *Journal of Psycholoeducational Assessment, 29*: 364. 63.

Walker, C. M., Beretvas, S. N., & Ackerman, T. A. (2001). An examination of conditioning variables used in computer adaptive testing for DIF. *Applied Measurement in Education*, *14*, 3-16.

Walker, E.R., Engelhard, G., & Thompson, N.J. (2012). Using Rasch measurement theory to assess three depression scales among adults with epilepsy. *Seizure*, *21*(6), 437-443. http://dx.doi.org/10.1016/j.seizure.2012.04.009

WHOQOL Group (2006). A cross-cultural study of spirituality, religion, and personal beliefs as components of quality of life. *Social Science & Medicine, 62,* 1486–1497. https://doi.org/10.1016/j.socscimed.2005.08.001

Windle, G., Bennett, K.M., & Noyes, J. (2011). A methodological review of resilience measurement scales. *Health and Quality of Life Outcomes*, *9*(1), 8.

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

Wright, B.D., & Linacre, J.M. (1994). Reasonable mean-square fit values. *Rasch Measurement Transactions*, *8*(3), 370-371. Retrieved from https://www.rasch.org/rmt/rmt83b.htm

Yaşar, M. & Aybek, E. C. (2019). Üniversite öğrencileri için bir yılmazlık ölçeğinin geliştirilmesi: madde tepki kuramı temelinde geçerlilik ve güvenilirlik çalışması. *İlköğretim Online*, *18*(4) 1687-1699. DOI: 10.17051/ilkonline.2019.635031

Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, *8*(2), 125-145.

Zanon, C., Hutz, C. S., Yoo, H., & Hambleton, R. K. (2016). An application of Item Response Theory to psychological test development. *Psicologia: Reflexao e Critica, 29*, 18. https://doi.org/10.1186/s41155-016-0040-x.

Zieky, M. (1993). DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 337–347). Erlbaum.

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.* Ottawa, Ontario, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.