Kalem Eğitim ve İnsan Bilimleri Dergisi 2024, 14(1), 241-261, doi: 10.23863/kalem.2024.282

A Study on the Use of the Rasch Model in the Analysis of Data Collected by Ranking Judgments[1]

Doç. Dr. Mustafa İLHAN

Dicle Üniversitesi, Ziya Gökalp Eğitim Fakültesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, Diyarbakır / Türkiye, mustafailhan21@gmail.com, ORCID: 0000-0003-1804-002X

Doç. Dr. Melek Gülşah ŞAHİN

Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Ankara / Türkiye, mgulsahsahin@gazi.edu.tr, ORCID: 0000-0001-5139-9777

Prof. Dr. Neşe GÜLER

İzmir Demokrasi Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Ankara / Türkiye, gnguler@gmail.com, ORCID: 0000-0002-2836-3132

Abstract

This research aimed to exemplify the use of the Rasch model in the analysis of data on ranking judgments. To this end, in the study, we delineated in detail all the processes related to the Rasch analysis, from the preparation of the data file to the interpretation of the analysis outputs. Besides, we compered the Rasch analysis results with the outputs obtained from traditional method. We conducted our study on 261 secondary school students. We collected the research data by employing an instrument

[1] A part of this study was orally presented at the 1st International Conference on Educational Research organized by İzmir Demokrasi University between 4-6 October 2019.

*Sorumlu Yazar. Tel: +90 412 241 10 00-8861 | Araştırma Makalesi.

Makale Tarih Bilgisi. Gönderim: 21.07.2022, Kabûl: 18.01.2023, Erken Görünüm: Nisan, 2024, Basım: Haziran, 2024

containing five stimuli regarding the positive aspects and limitations of the mathematics homework. We asked students to rank the positive aspects of the homework by coding the most prominent advantage with 1 and the last advantage with 5. Similarly, we requested them to put in order the limitations of the homework by coding the foremost limitation with 1 and the last limitation with 5. We analyzed the research data according to both the Rasch model and traditional method. Research results revealed that the agreement between the stimuli's scale values obtained from the Rasch analysis and from the traditional method was high. Considering this result and that the Rasch model is a more practical option in terms of the analysis process, we can suggested researchers to prefer Rasch model instead of the traditional method in the analysis of data based on ranking judgments.
Keywords: Scaling approaches; Stimulus-centered scaling; Ranking judgements; Ranking of stimuli; Rasch model.

Sıralama Yargılarıyla Toplanan Verilerin Analizinde Rasch Modelinin Kullanımına Yönelik Örnek Bir Çalışma

Öz

Bu araştırma, sıralama yargılarıyla toplanan verilerin analizinde Rasch modelinin kullanımını örneklendirmek amacıyla yapılmıştır. Bu amaçla çalışmada, veri dosyasının hazırlanmasından analiz çıktılarının yorumlanmasına kadar Rasch analizi ile ilgili tüm süreçler ayrıntılı olarak açıklanmıştır. Ayrıca çalışma kapsamında Rasch analizi sonuçları geleneksel yöntemden elde edilen çıktılarla karşılaştırılmıştır. Çalışma 261 ortaokul öğrencisi üzerinde yürütülmüştür. Araştırma verileri, matematik ödevinin olumlu yönleri ve sınırlılıkları ile ilgili beşer uyarıcı içeren bir ölçme aracı kullanılarak toplanmıştır. Öğrencilerden matematik ödevlerinin en önemli avantajını 1 ve en sonda gelen avantajını 5 ile kodlayarak ödevin olumlu yönlerini sıralamaları istenmiştir. Benzer şekilde, en önemli sınırlılığı 1 ve en sonda gelen sınırlılığı 5 ile kodlayarak matematik ödevlerine ilişkin sınırlılıkları sıralamaları talep edilmiştir. Araştırma verileri hem Rasch modeline hem de geleneksel yönteme göre analiz edilmiştir. Araştırma sonuçları, uyarıcılar için Rasch analizinden ve geleneksel yöntemden elde edilen ölçek değerleri arasında yüksek bir uyum bulunduğunu ortaya koymuştur. Bu sonuç ve analize ilişkin süreçler açısından Rasch modelinin daha pratik bir seçenek olduğu göz önüne alındığında sıralama yargılarıyla toplanan verilerin analizinde geleneksel yöntem yerine Rasch modelinin tercih edilmesi önerilebilir.

Anahtar Kelimeler: Ölçekleme yaklaşımları; Uyarıcı-merkezli ölçekleme; Sıralama yargıları; Uyarıcıların sıralanması; Rasch modeli.

Introduction

Most of the traits that are the subject of behavioral, social and educational sciences cannot be observed directly. For example, the constructs such as intelligence, personality, interest, motivation, attitude, anxiety, and self-efficacy cannot be observed directly and their physical size is unknown. Therefore, it is rather difficult to measure these kinds of variables when compared to physical properties. However, in order to shed light on human behaviors and the cause-effect relationships between behaviors, it is extremely important to make these structures, which are known to exist but cannot be directly observed, measurable (Kan, 2008a). Attempts to find methods that can be used to measure psychological variables that are unknown how to measure and do not have standard measurement tools have laid the groundwork for scaling studies (Turgut and Baykul, 1992).

Scaling studies are interested in finding connections between the measured quantities of physical stimuli and their perceived quantities (Erkuş, 2003). Price (2017) defines the scaling term as the process of measuring stimuli by way of a mathematical representation of the stimulus–response curve, based on the explanations of Birnbaum (1998), Guilford (1954) and Torgerson (1958). A scaling model provides an operational or a relational framework for attaining scores (or numerical categories) on a construct obtained from a series of individuals, objects, or events, thereby facilitating the transformation from qualitative constructs into measurable metrics (Price, 2017). Scaling models may be used to scale individuals, stimuli and or both individuals and stimuli (McIver and Carmines, 1981). Hence, scaling models are categorized under three headings: stimulus-centered, subject centered and response centered (Torgerson, 1958). These three models of scaling are summarized as in Table 1.

Table 1 implies that the source of variation in scale scores is different in each scaling approach. In stimuli-centered scaling systematic response variations (by respondents) is ascribed to differences among referent stimuli (Dobrzykowski, 1998). In subject-centered approach, the source of variation across scale scores is deemed as individuals. In the response-centered scaling, on the other hand, the variation in scale scores is attributed to both stimuli and individuals (Mokken, 1971). In addition, as can be seen in Table 1, there are various techniques that can be used in each scaling approach. Nevertheless, this research will center upon only ranking judgments, which is one of the stimuli-centered scaling approaches.

Table 1. Three Approaches to Scaling*

Method	Purpose	Instances
Stimulus-centered	Positioning stimuli or items on a continuum.	A group of people who are requested to evaluate the weight of a given set of objects. In this example, objects are quantified with respect to a designated attribute (e.g., pairwise comparisons, ranking judgments, Osgood's semantic differential scale)
Subject-centered	Locating individuals at different points on a continuum.	Subjects are asked to indicate their degree of agreement or disagreement to a series of statements about a single topic. The scale score of a subject is calculated using his/her responses (e.g. Likert scales, Stephenson's Q-sort technique).
Response-centered	Placing both subjects and stimuli on a common continuum. Response data are used to scale subjects on a continuum, based on the strength of the stimuli/items endorsed (or answered correctly); simultaneously, stimuli are scaled in terms of the strength or amount of the trait possessed by the subjects who endorse them.	A patient is asked to select from a set of health-state descriptions the state that is most alike to his/her own health condition (e.g., Guttman scaling, Coomb's unfolding model, item response theory, latent class analysis, and mixture models).

* *The table was created with reference to Shin (1974), McIver and Carmines (1981), Crocker and Algina (1986), Smith and Albaum (2005), Zhu and Yang (2016), Krabbe (2017) and Price (2017).*

Ranking Judgments

This scaling technique is relying on the judge(s) ranking a series of stimuli presented to them in order of perceived size/priority. To accomplish this, *m* numbers of stimuli are presented to individual(s) and they are judged repeatedly by the same individual, or independently by the multiple individuals (Briggs, 2022). Individual(s) are asked to consider all the stimuli, to compare each stimulus with the others and thuswise assign a sequence number to the stimuli (Kan, 2008b). Since this method compels the judges to make the greatest possible distinction between stimuli, it produces results with high validity where the judge is able to make this distinction. Ranking judgments are similar to pairwise comparisons technique in terms of revealing the magnitude-smallness relationships between stimuli. On the other hand, unlike the pairwise comparisons, in ranking judgments stimuli are presented to the judges all in one, not in paired. So, this method is faster than pairwise comparisons and is usually easier and more motivating to the respondents (Krishnaswamy, 2006). Besides, ranking judgements averts inconsistent triples that threaten internal consistency in pairwise comparisons (Turgut and Baykul, 1992).

Analysis of Data Collected by Ranking Judgments

When we reviewed the literature, we realized that the analyses were carried out in Microsoft Excel on the grounds the traditional method in studies in which the scaling with ranking judgments was utilized (e.g., Bal, 2011; Bozgeyikli and Kesici, 2016; Bozgeyikli and Toprak, 2013; Demirçelik et al., 2021; Gezer and İlhan, 2018; Örs Özdil and Kınay, 2015; Özdemir, 2021; Şahin et al., 2015; Şahin, Sarkın and Taşdemir, 2019; Toprak et al. 2020; Yalçın and Şengül-Avşar, 2014; Yıldırım, Seheryeli and Anıl, 2020). A multi-step process is followed while the data collected by means of the ranking judgments are analyzed according to traditional method in Microsoft Excel. Primarily, rank frequencies are generated that show how many times each stimulus in the measurement tool is placed in which sequence. In the second step, the frequencies matrix is calculated by using rank frequencies. Following, the ratio matrix is obtained from the frequency matrix. Then, the z scores corresponding to the values in the cells of the ratio matrix are found and thus the unit normal deviations matrix is obtained. In the last step, the sum of the values of each column is added to the bottom row of the unit normal deviations matrix and the scale values are reached by dividing the values in this row by the number of stimuli. Testing the reliability of the scale values acquired requires the researcher to perform some other operations. To this end; the researcher must construct the expected unit normal deviations matrix, the expected ratio matrix and the error matrix achieved by taking the absolute value of the difference between the cell values of the observed and expected ratio matrix. Subsequently he/she must determine the mean error amount through the errors matrix and check the significance of calculated mean error (Turgut and Baykul, 1992).

The above listed analysis procedures can be challenging or at least tedious drudgery for researchers. Conversely, we consider that the Rasch model will be a practical alternative that can be used in the analysis of data collected with ranking judgments. Because when the Rasch model is employed, the data on the ranking judgments can be simply analyzed without the need for consecutive processing steps, and the statistics that provide evidence for its validity and reliability are reported simultaneously with the scale values of the stimuli.

Existing Studies in the Literature

In the literature, there are many researches in which the Rasch model is used in the analysis of ordinal data. The most typical examples of this are the studies in which the Likert-type scale data are analyzed with the Rasch model (e.g., Alisat and Riemer, 2015; Anshel et al., 2009; Bonino et al., 2018; Brinthaupt and Kang, 2014; Behizadeha and Engelhard, 2014; Hopkins et al., 2021; İlhan and Güler, 2018; Ricketts et al., 2017; Şahin et al., 2022; Şen and Göçen, 2021). Despite the fact the data obtained via Likert scales are assumed to be interval, in reality, the distances between the response options are not equal in these scales, and therefore the data obtained remains at the ordinal-level techincally. Nonetheless, the results of Likert scales can be converted to the interval scale by means of Rasch analysis (Wright and Masters, 1982).

Also, there are studies in the literature in which the Rasch model is used for scaling and the results of Rasch analysis are compared with the outputs obtained from traditional analysis. Andrich (1978) theoretically discussed the relationships between the Thurstone and Rasch approaches to item scaling. Similarly, Jansen (1984) theoretically examined the relationships between the Thurstone, Coombs, and Rasch approaches to item scaling. Engelhard (1984), on the other hand, both theoretically and empirically described and compared Thorndike, Thurstone, and Rasch for calibrating test items. İlhan et al. (2021) operated the Rasch model when analyzing pairwise comparison data. However, we did not find any study in the literature in which the Rasch model was used in the analysis of the data collected by ranking judgments.

Purpose and Originality of the Research

Since we have not encountered a study in the literature in which the Rasch model was used in the analysis of the data collected with ranking judgments, we believe that an exemplary study on this subject will contribute to the literature. In this respect, the current research intends to depicture in detail the processes of analyzing data on ranking judgments via the Rasch model, from the preparation of the data file to the interpretation of the analysis outputs. Moreover, it purposes to test the agreement between the scale values obtained from the Rasch analysis and from the traditional method for the stimuli subjected to ranking. It is hoped that the paper will be help of to the researchers who intend to study based on ranking judgments.

Method

Research Model

As the research intends comparing the scale values obtained by analyzing the ranking judgements data through two different methods, the model of the study is the method-comparison. Method-comparison studies are aimed to examine the compatibility between the results procured from different methods or to test the relationships between the results attained in a relatively less used method and those achieved from the usual one (Hanneman, 2008).

Study Group

In the study, we did not aim to generalize about the universe, since we attempt to exemplify the use of the Rasch model in the analysis of the data collected according to the ranking judgments. We collected data from 261 secondary school students, 146 female and 115 male. Of the students, 225 were attending the eighth grade and 36 were attending the seventh grade.

Instrument and Data Collection Process

We developed the data collection tool used in the study ourselves. At the beginning of the data collection tool, we included the demographics of gender and grade level. Following these variables, we presented five stimuli for the positive aspects and limitations of the homework given in the mathematics course. We identified the stimuli putting account the relevant literature (Aladağ and Doğu, 2009; Gedik et al., 2011; Sarıgöz, 2011; Yücel, 2004). After we prepared the draft form for the instrument, we got opinions from two experts from the field of mathematics education, three from the field of measurement and evaluation, and one expert from the field of Turkish language.

All experts stated that the instument contains the main positive aspects and limitations of the homeworks that can be perceived by the students, but some of them proposed several changes in some stimuli. For example, we revised the stimulus “Increases my interest in the relevant subject”, which we gave place among the positive aspects of the homework, as “Makes me enjoy the relevant subject more” in line with opinions. Similarly, we changed the statement “Makes it easier for me to relate what I’ve learned to daily life” to “Allows me to make connections between what I have learned in the course and daily life”. Lastly, we altered the stiumulus “I'm having trouble accessing the resources I need to use in homework" to "When I need to use a resource (internet, book, encyclopedia, etc.) for the given homework, I have trouble reaching this resource". Thus we finalized the data collection tool. Table 2 show the stimuli presented to the students in the instrument.

Table 2. Stimuli Presented to the Students Regarding the Positive Aspects and Limitations of Homework

Stimuli on the Positive Aspects of Homework	Helps me to repeat what I learned in the lesson.
	Helps me come prepared to the next lesson.
	Makes me enjoy the relevant subject more.
	Assists me prepare for exams
	Allows me to make connections between what I have learned in the course and daily life.
Stimuli on the limitations of Homework	Takes a lot of time to do my homework
	When I need to use a resource (internet, book, encyclopedia, etc.) for the given homework, I have trouble reaching this resource.
	I have difficulty understanding the content of the homework given
	Teachers give too much homework
	The homework that are given require me to get help from someone.

We asked students to rank the positive aspects of the homework by coding the most prominent advantage with 1 and the last advantage with 5. Similarly, we requested them to put in order the limitations of the homework by coding the foremost limitation with 1 and the last limitation with 5. We collected the data in students' own classrooms on the voluntary basis in September 2019.

Data Analysis

We analyzed the research data both according to the Rasch model and traditional method. Rasch analysis was run through FACETS package program. There were two facets in Rasch analysis: students and stimuli. However, we defined the student facet as "anchor" and kept the students' measurements constant at 0 logit. Also, we defined the stimulus facet as *negative-oriented* in order to specify that the stimuli with a low number were seen more prioritized by the participants (see Linacre, 2022, p.58-59). Figure 1 denotes the syntax we used to analyze the data we collected regarding the positive sides of the homework. We carried out, on the other hand, the analysis based on traditional method using Microsoft Excel. We used Kendall Tau correlation (τ) to determine the consistency between the traditional method and the scale values of Rasch analysis as it generates more accurate results when the number of observations is less than 10 (Hahs-Vaughn and Lomax, 2020).

```
Title = POSITIVE ASPECTS OF HOMEWORKS
Facets=2
Positive=0
Arrange=m,2N
Non-centered=1
Unexpected=4
Usort=1,2,3
Barcharts=No
Vertical=1*,2N
Yardstick=0,10
Zscore=,2

Models=?,#,R5
Labels=
1=STUDENTS,A
1-261
*
2=STIMULANTS
1-5
*
Data=
1       1-5     1       3       4       2       5
2       1-5     1       2       4       3       5
.       .       .       .       .       .       .
.       .       .       .       .       .       .
.       .       .       .       .       .       .
.       .       .       .       .       .       .
.       .       .       .       .       .       .
.       .       .       .       .       .       .
.       .       .       .       .       .       .
```

Figure 1. The Syntax Used to Analyze The Data Regarding The Positive Aspects of The Homework

Results

Figure 2 demonstrates the variable map obtained by analyzing the data on the positive aspects of the homework. The leftmost column of the variable map comprises the measurement levels. The range in which measures are reported in this column depends on the measurements of the components on the facets included in the study and therefore varies from one research to another. When we examined the Figure 2 we saw that the measures for the current study are reported to be in ± 1 range.

The second column of the variable map contains the measures of the student facet. In studies that aim to measure the ability levels of students (i.e., in response-centered or subject-centered approaches), it can be seen how the examinees are ranked in terms of their ability levels by looking at this column. Nevertheless, since ranking judgments are a stimuli-centered technique, it is out of the question of obtaining an ability score for students. For this reason, we defined student facet as anchor and kept it constant at 0 logit. Consequently, the measures of this facet were clustered at 0 point of the variable map.

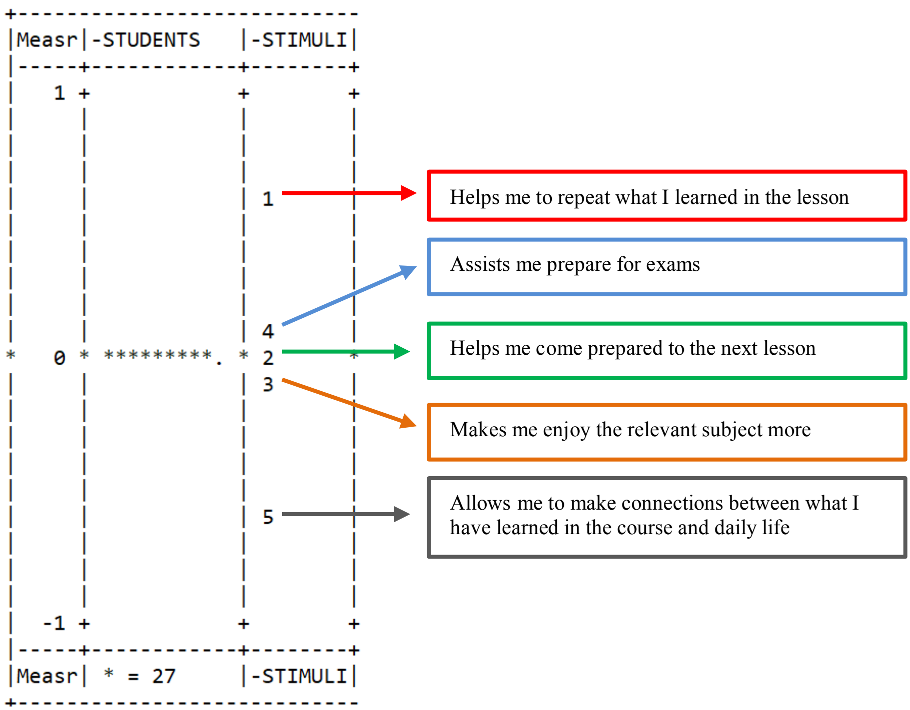

Figure 2. The variable map for the positive aspects of the homework

In the third column of the variable map, there are the stimuli. Throughout this column, there is a top-down order from the most important advantage to the relatively less important advantage of homework. Accordingly, Figure 2 indicates that the most important advantage of homework for students is that it helps them to repeat what they learned in the lesson. It was seen by the students as the last among the advantages of homework that it allowed making connections between what was learned in the course and daily life. In addition, when we look at Figure 2, it is striking that the stimuli with the encodings of two, three and four are located close to each other. The ranking mentioned can also be followed from the measurement reports in Table 3.

Table 3. Measurement Report of the Stimulus Facet for the Positive Aspects of the Homework

Stimuli	Measure	Infit MnSq	Outfit MnSq
1	.55	1.00	1.00
2	.04	1.00	1.00
3	–.12	1.00	1.00
4	.15	1.00	1.00
5	–.62	1.00	1.00
Reliability: .98		Separation: 8.07	
X^2_4 = 272.70, $p< .001$			

Table 3 reflects that the measure of the 1-coded stimulus is significantly higher compared to the other stimuli. The measure of the 5-coded stimulus is significantly lower the others. The scale values of the other three stimuli, on the other hand, are close to each other. The fact that infit and outfit mean-square values in Table 3 are equal to 1.00, which is the expected value of these statistics, means that the model-data fit was achieved. Furthermore, the reliability coefficient and the separation ratio were estimated as .98 and 8.07, respectively. The reliability index obtained from the Rasch analysis bounded by 0 and 1 analogous to Cronbach' alpha coefficient (Bond et al., 2012). The separation ratio, on the other hand, ranges from 1 to ∞ (Sudweeks et al., 2004). Low values for separation (<3) and reliability (<.90) of the item facet hint that the person sample is not large enough to confirm the item difficulty hierarchy of the instrument (Linacre, 2012). Considering these criteria, we can assert that our sample is large enough and the scale values calculated for the stimuli are reliable. The fact that the chi-square value in the table was significant also signifies that the stimuli could be distinguished effectively from each other by the students. After analyzing the data on the positive aspects of homework, we checked up on the data on the limitations. Figure 3 illustrates the variable map reported when we analyzed rankings done by students regarding the limitations of homework.

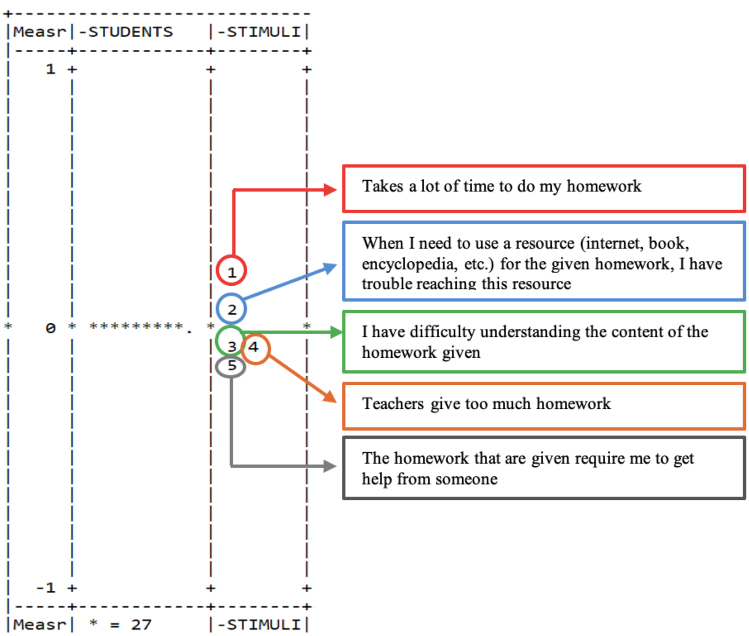

Figure 3. The variable map for the limitations of the homework

Figure 3 displays that the stimuli regarding the limitations of the homework clustered in a narrower range than the stimuli related to its positive sides. The stimulus “it takes a lot of time to do my homework” was seen by the students as the most limitation side of the homework. “The homework that are given require me to get help from someone” was ranked last among the stimulus regarding the limitations of the homework. This ranking can also be seen in the measurement reports in Table 4.

Table 4. Measurement Report of The Stimulus Facet for The Limitations of The Homework

Stimuli	Measure	Infit MnSq	Outfit MnSq
1	.19	1.00	1.00
2	.10	1.00	1.00
3	–.06	1.00	1.00
4	–.11	1.00	1.00
5	–.12	1.00	1.00
Reliability: .90		Separation: 2.92	
X_4^2= 39.60, p<.001			

Parallel to the fact that the measures of the stimuli related to the limitations of the homework are in a narrow range, the reliability coefficient, separation ratio and chi-square value in Table 3 were lower than those calculated for the stimuli on the positive aspects of the homework. Namely, students could not distinguish the stimuli concerning the limitations of the homework as effectively as the stimuli on its positive aspects. After the Rasch analysis was completed, the ranking judgments data were analyzed with the traditional method. Table 5 and Table 6 shows the matrices concerning the traditional analysis.

Table 5. Results Regarding the Analysis Via the Traditional Method for the Positive Aspects of Homework

		Positive Aspects of the Homework				
		S1	S2	S3	S4	S5
Matrix of rank frequencies	1	147	23	31	46	147
	2	58	60	52	76	58
	3	19	101	43	67	19
	4	22	59	88	48	22
	5	15	18	47	24	15
Matrix of frequencies	S1					
	S2	52343				
	S3	53248.5	39485			
	S3	48492.5	29797.5	25906.5		
	S5	60458	54052.5	49079	54867.5	

Ratio matrix	S1					
	S2	0.7684				
	S3	0.7817	0.5796			
	S3	0.7119	0.4374	0.3803		
	S5	0.8875	0.7935	0.7205	0.8054	
Unit normal devitions matrix	S1					
	S2	0.734				
	S3	0.778	0.201			
	S4	0.559	-0.158	-0.305		
	S5	1.213	0.819	0.584	0.861	
$\sum z_J$		3.284	0.128	-0.699	0.765	-3.477
Z_{Jort}		0.657	0.026	-0.140	0.153	-0.695
S_J		1.352	0.721	0.555	0.848	0.000

Table 6. Results Regarding the Analysis Via the Traditional Method for the Limitations of Homework

		Limitations of the Homework				
		S1	S2	S3	S4	S5
Matrix of rank frequencies	1	83	66	33	42	37
	2	61	49	55	52	44
	3	49	39	77	49	47
	4	28	60	52	51	70
	5	40	47	44	67	63
Matrix of frequencies	S1					
	S2	38619				
	S3	41954.5	36441			
	S3	42605.5	37979	35791.5		
	S5	34179.5	39367	37645	35466	
Ratio matrix	S1					
	S2	0.5669				
	S3	0.6159	0.5349			
	S3	0.6254	0.5575	0.5254		
	S5	0.5017	0.5779	0.5526	0.5206	
Unit normal devitions matrix	S1					
	S2	0.169				
	S3	0.295	0.088			
	S4	0.320	0.145	0.064		
	S5	0.004	0.197	0.132	0.052	
$\sum z_J$		0.787	0.260	-0.186	-0.476	-0.385
Z_{Jort}		0.157	0.052	-0.037	-0.095	-0.077
S_J		0.252	0.147	0.058	0.000	0.018

The last rows of Table 5 and Table 6, labeled *Sj*, contains the scale values calculated for the stimuli. In these rows, the fact that one stimulus has a higher scale value than the others is interpreted as the relevant stimulus being seen more primarily by the judges. Eventually, we checked the agreement between the scale values calculated for the stimuli in the traditional method and in the Rasch analysis. Table 7 exhibits the outputs of this comparison.

Table 7. Agreement Between the Stimuli's Scale Values in the Traditional Method and in the Rasch Analysis

	Positive Aspects of the Homework				Limitations of the Homework			
	Rasch Analysis		Traditional Method		Rasch Analysis		Traditional Method	
	Scale Value	Ranking	Scale Value	Ranking	Scale Value	Ranking	Scale Value	Ranking
S1	.55	1	1.352	1	.19	1	.252	1
S2	.04	3	.721	3	.10	2	.147	2
S3	–.12	4	.555	4	–.06	3	.058	3
S4	.15	2	.848	2	–.11	4	.000	5
S5	–.62	5	.000	5	–.12	5	.018	4
	$\tau = 1.00$				$\tau = .80$			

Table 7 displays that there is an exact agreement between the scale values calculated in the Rasch analysis and the traditional method for the stimuli of the positive aspects of the homework (τ=1.00). For the limitations of the homework the agreement between the scale values calculated in the two methods is quite strong, if not perfect (τ=.80). These results infer that the two methods generally effectuate similar results.

Discussion, Conclusion and Future Directions

Rasch analysis is a statistical technique constructing interval measures from raw ordinal observations (Granger et al., 1993). Correspondingly, it is one of the models that can be employed for the analysis of data based on ranking judgments. This research exemplified the use of the Rasch model in the analysis of data collected relying on ranking judgments. We utilized the rankings made by the students regarding the positive aspects and limitations of the mathematics homework as a data source. The results obtained both Rasch analysis and traditional method revealed that the most positive feature of math homework is that it helps students to repeat what they have learned in the lesson. The stimulus of "Takes a lot of time to do my homework" was remarked as the foremost limitation of math homework by students. What is striking about the research results is the agreement between the stimuli's scale values derived from the Rasch analysis and from the traditional method was complete for the positive aspects of the homework and strong for the limitations of the homework. This result accords with many studies (e.g., Awopeju and Afolabi, 2016; Courville, 2004; Fan, 1998; Hwang, 2002; Kan, 2006; Progar and Sočan, 2008; Uysal, 2015) that found different measurement theories produce similar item parameters.

The results reported in studies comparing the Rasch model with different scaling approaches are in line with those reached in our study. For example, Andrich (1978) and Jansen (1984) compared Thurstone's Case V equation and Rasch model theoretically in a mathematical framework and they stated that the parameters calculated in the two methods would be the identical. Likewise, Guler, İlhan and Taşdelen-Teker (2018) specified that both pairwise comparisons scaling technique and Rasch analysis can be used to bring the ordinal data to the interval in their study, and that the scale values obtained by these two methods will be consistent.

When we interpret the results of our research in the light of the listed studies in the literature, we can enounce the followings: Considering the substantial agreement between the two procedures and the fact that the Rasch model is a more practical option for analysis processes, we we can suggested researchers seeking answers to the question of how to most easily analyze data based on ranking judgments to prefer Rasch model instead of the traditional method. Indeed, Güzeller et al. (2016) remarked that scaling analyzes carried out in Microsoft Excel are time-consuming and error-prone. In the same vein, Tat and Anıl (2016) expressed that the scaling analyzes conducted in statistical softwares are more advantageous than the analyzes performed in Microsoft excel when it comes to two methods that are known to produce similar results. Therewithal, as the results acquired will be similar no matter which method is used, the researchers who feel more competent in the traditional one can continue to use this method.

Briefly; data collection process are simple in stimulus-centered scaling approaches such as pairwise comparisons and ranking judgements, but the analysis of the data collected via these techniques can be difficult for some researchers. Johnson and Christensen (2014) also adverted to this situation and said that rank order items are difficult to statistically analyze and relate to other variables. Moreover; as reliability calculations require additional steps in the traditional method, ranking judgments data are usually used without reports psychometric properties. Analyzes based on the Rasch model can be a functional alternative that will serve to overcome this difficulties and restrictions. Based on this consideration, İlhan et al. (2021) employed Rasch model in the analysis of the data collected according to the pairwise comparison technique. By contrast with, no previous study which used Rasch model in the analysis of data on ranking judgements was found in the literature. In this regard, we believe that this study has the potential to contribute to researchers. However, whether the rankings of the stimuli vary according to the certain independent variables was not addressed in this study. Hence, a natural progression of current study is the inclusion of various independent variable(s) as a facet in the Rasch analysis besides to the stimulus and individual/person facets for testing the effect of these independent variable(s) on the stimuli's measures.

Another suggestion for further research may be utilizing different statistical softwares when analyzing ranking judgments data according to the Rasch model. In current research, we performed Rasch analyzes in the FACETS program developed by Linacre (1988) for many-facet Rasch analysis. The same analyzes can also be implemented via WINSTEPS package program (Linacre, 2006). In addition, since the many-facet Rasch analyzes performed in the FACETS program can also be done in the *TAM (Test anaylsis modules)* package of R software (see Wind and Hua, 2022), the analysis of the ranking data can also be tried in the TAM package in further research. Finally, the *prm* (Lee and Yu, 2013) and *PlackettLuce* (Turner et al., 2020) packages of may be another alternatives that researchers can apply to analyze ranking data.

References

Aladağ, C. and Doğu, S. (2009). The evaluation of homeworks which are given at secondary school according to students views. *Selçuk University Social Sciences Institute Journal, 21,* 15-23. https://dergipark.org.tr/tr/pub/susbed/issue/61797/924362

Alisat, S. and Riemer, M. (2015). The environmental action scale: Development and psychometric evaluation. *Journal of Environmental Psychology*, *43*, 13-23. Doi: 10.1016/j.jenvp.2015.05.006

Andrich, D. (1978). Relationships between the Thurstone and Rasch approaches to item scaling. *Applied Psychological Measurement*, *2*(3), 449-460. Doi: 10.1177/014662167800200319

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

Awopeju, O. A. and Afolabi, E. R. I. (2016). Comparative analysis of classical test theory and item response theory based item parameter estimates of senior school certificate mathematics examination. *European Scientific Journal*, *12*(28), 263-284.
Doi: 10.19044/esj.2016.v12n28p263

Bal, Ö. (2011). Scaling of the factors that were considered to be effective in the success in placement exam [Seviye Belirleme Sınavı (SBS) başarısında etkili olduğu düşünülen faktörlerin sıralama yargıları kanunuyla ölçeklenmesi]. *Journal of Measurement and Evaluation in Education and Psychology, 2*(2), 200-209. https://dergipark.org.tr/tr/pub/epod/issue/5804/77230

Behizadeha, N. and Engelhard, G. (2014). Development and validation of a scale to measure perceived authenticity in writing. *Assessing Writing*, *21*, 18-36.
Doi: 10.1016/j.asw.2014.02.001

Birnbaum, M. H. (Ed.). (1998). *Measurement, judgment, and decision making* (2nd ed.). San Diego, CA: Academic.

Bond, T. G., Yan, Z. and Heene, M. (2012). *Applying the Rasch model: Fundamental measurement in the Human Sciences* (4th ed.). New York: Routledge.

Bonino, S., Graziano, F., Borghi, M., Marengo, D., Molinengo, G. and Calandri, E. (2018). The self-efficacy in multiple sclerosis (SEMS) Scale: Development and validation with Rasch analysis. *European Journal of Psychological Assessment*, *34*(5), 352-360.
Doi: 10.1027/1015-5759/a000350

Bozgeyikli, H. and Kesici, Ş. (2016). Scaling undergraduate's leisure activities according to ranking judgments. *Journal of Youth Research*, *4*(1), 39-72.

Bozgeyikli, H. and Toprak, E. (2013). University youth's mate selection criteria by rank order judgments scaling. *Journal of Youth Research*, *1*(1), 68-87.

Briggs, D. C. (2022). A history of scaling and its relationship to measurement. In B. E. Clauser and M. B. Bunch, (Eds.), *The history of educational measurement: Key advancements in theory, policy and practice* (263-291). New York: Routledge.
Doi: 10.4324/9780367815318

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

Courville, T. G. (2004). *An empirical comparison of item response theory and classical test theory item/person statistics*. Unpublished doctoral dissertation, Texas A&M University.
https://core.ac.uk/download/pdf/147123147.pdf

Crocker, L. and Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart, & Winston.

Demirçelik, E., Ilgın Bağcı, M. and Usta, İ. (2021). Scaling the problems of gifted high school students during the pandemic process by rank-order judgments. *International Journal of Social Sciences in Turkish Cultural Geography, 6*(2), 277-289.
https://dergipark.org.tr/en/pub/turksosbilder/issue/67694/973089

Dobrzykowski, E. A. (1998). *Essential readings in rehabilitation outcomes measurement: Application, methodology, and technology*. Gaithersburg, MD: Aspen.

Engelhard, G. (1984). Thorndike, Thurstone, and Rasch: A comparison of their methods of scaling psychological and educational tests. *Applied Psychological Measurement*, *8*(1), 21-38. Doi: 10.1177/014662168400800104

Erkuş, A. (2003). *Essays on psychometry [Psikometri üzerine yazılar].* Ankara: Turkish Psychologists Association.

Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person parameters. *Educational and Psychological Measurement*, *58*, 357-381. Doi: 10.1177/0013164498058003001

Gedik, N., Altıntaş, E. and Kaya, H. (2011). Students' views about homework in Science and technology course. *Journal of European Education*, *1*(1), 6-13. http://www.eu-journal.org/index.php/JEE/article/view/170/156

Gezer, M. and İlhan, M. (2018). An investigation of the factors causing inequality of opportunity in education in Turkey through rank order judgement from the perspective of academicians. *Journal of Higher Education*, *8*(3), 301-312. Doi: 10.2399/yod.18.016

Granger, C. V., Hamilton, B. B., Linacre, J. M., Heinemann, A. W. and Wright, B. D. (1993). Performance profiles of the functional independence measure. *American Journal of Physical Medicine & Rehabilitation*, *72*(2), 84-89. Doi: 10.1097/00002060-199304000-00005

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York: McGraw-Hill.

Güler, N., İlhan, M. and Taşdelen-Teker, G. (2018). Comparing the scale values obtained from pairwise comparison scaling method and Rasch analysis. *Inonu University Journal of the Faculty of Education*, *19*(1), 31-48. Doi: 10.17679/inuefd.400386

Güzeller, C. O., Eser, M. T. and Aksu, G. (2016). Pair-wise comparasion method application via R project and microsoft excel. *Journal of Measurement and Evaluation in Education and Psychology*, *7*(1), 96-108. https://dergipark.org.tr/tr/pub/epod/issue/27272/287103

Hahs-Vaughn, D. L. and Lomax, R. G. (2020). *Statistical concepts–A first course*. New York: Taylor & Francis. Doi: 10.4324/9780429261268

Hanneman, S. K. (2008). Design, analysis, and interpretation of method-comparison studies. *AACN Advanced Critical Care*, *19*(2), 223-234. Doi: 10.1097/01.AACN.0000318125.41512.a3

Hopkins, E. G., Lyndon, M. P., Henning, M. A. and Medvedev, O. N. (2021). Applying Rasch analysis to evaluate and enhance the academic motivation scale. *Australian Journal of Psychology*, *73*(3), 348-356. Doi: 10.1080/00049530.2021.1904794

Hwang, D. Y. (2002, February). *Classical test theory and item response theory: Analytical and empirical comparisons*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX. https://files.eric.ed.gov/fulltext/ED466779.pdf

İlhan, M. and Güler, N. (2018). The use of Rasch model in likert types scales: An application on the fear of negative evaluation scale-student form (FNE–SF). *Trakya Journal of Education*, *8*(4), 756-775. https://dergipark.org.tr/tr/pub/trkefd/issue/39371/357367

İlhan, M., Güler, N. and Taşdelen-Teker, G. (2021). Views of teachers on the potential negative effects of high stake tests. *International Journal of Assessment Tools in Education*, *8*(2), 394-408. Doi: 10.21449/ijate.735232

Jansen, P. G. W. (1984). Relationships between the Thurstone, Coombs, and Rasch approaches to item scaling. *Applied Psychological Measurement*, *8*(4), 373-383. Doi: 10.1177/014662168400800401

Johnson, B. and Christensen, L. (2014). *Educational research: Quantitative, qualitative, and mixed approaches* (5th ed.). Thousand Oaks, California: Sage.

Kan, A. (2006). An empirical study on the comparison of predicted item parameters with respect to classical and item response test theories. *Mersin University Journal of the Faculty of Education*, *2*(2), 227-235. https://dergipark.org.tr/tr/download/article-file/160995

Kan, A. (2008a). A comparison between scaling procedures for measuring psychological variables. *Journal of Theory and Practice in Education*, *4*(1), 2-18. https://dergipark.org.tr/tr/pub/eku/issue/5447/73868

Kan, A. (2008b). A comparison of scaling methods based on judge decisions: an empirical study. *Hacettepe University Journal of Education*, *35*, 186-194. http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/552-published.pdf

Krabbe, P. F. M. (2017). *The measurement of health and health status concepts, methods and applications from a multidisciplinary perspective*. London: Elsevier.

Krishnaswamy, K. N., Sivakumar, A. I. and Mathirajan, M. (2006). *Management research methodology: integration of principles, methods and techniques*. New Delhi: Pearson Education.

Lee, P. H. and Yu, P. L. H. (2013). An R package for analyzing and modeling ranking data. *BMC Medical Research Methodology*, *13*, 65. Doi: 10.1186/1471-2288-13-65

Linacre, J. M. (1988). *Facets: A computer program for many facet Rasch measurement*. Chicago: MESA. https://www.winsteps.com/a/Linacre-MFRM-book.pdf

Linacre, J. M. (2006). Rasch analysis of rank-ordered data. *Journal of Applied Measurement*, *7*(1), 129-139. https://www.winsteps.com/a/Linacre-rank-order-JAM.pdf

Linacre, J. M. (2012). *Winsteps Rasch measurement computer program user's guide*. https://www.winsteps.com/winman/reliability.htm

Linacre, J. M. (2023). *A user's guide to FACETS Rasch-model computer programs*. https://www.winsteps.com/a/Facets-Manual.pdf

McIver, J. P. and Carmines, E. G. (1981). *Unidimensional scaling*. Thousand Oaks, CA: Sage. Doi: 10.4135/9781412986441

Mokken, R. J. (1971). *A theory and procedure of scale analysis: With applications in political research*. The Hague, Paris: Mouton.

Örs-Özdil, S. and Kınay, E. (2015). Scaling 5th grade elective course preferences with rank-order judgments. *Journal of Measurement and Evaluation in Education and Psychology, 6*(2), 268-278. Doi: 10.21031/epod.30416

Özdemir, G. (2021). Using rank-order judgments scaling for political tactics used by teachers and school principals. *Eurasian Journal of Educational Research*, *93*, 1-18.
Doi: 10.14689/ejer.2021.93.1

Price, L. R. (2017). *Psychometric methods: Theory into practice*. New York: Guilford.

Progar, Š. and Sočan, G. (2008). An empirical comparison of item response theory and classical test theory. *Horizons of Psychology*, *17*(3), 5-24.
http://psiholoska-obzorja.si/arhiv_clanki/2008_3/progar_socan.pdf

Ricketts, S. N., Engelhard, G. and Chang, M. L. (2017). Development and validation of a scale to measure academic resilience in mathematics. *European Journal of Psychological Assessment, 33*(2), 79-86.
Doi: 10.1027/1015-5759/a000274

Sarıgöz, O. (2011). Assessment of the thoughts of secondary education students about homeworks given in chemistry lesson. *Electronic Journal of Vocational Colleges*, *1*(1), 80-87.
https://dergipark.org.tr/en/pub/ejovoc/issue/5395/73179

Shin, A. M. (1974). Relations between scales. In H. M. Blalock, (Ed.), *Measurement in the social sciences: theories and strategies* (121-158). Chicago: Aldine.

Smith, S. M. and Albaum, G. S. (2005). *Fundamentals of marketing research*. Thousand Oaks, California: Sage.

Sudweeks, R. R., Reeve, S. and Bradshaw, W. S. (2004). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing*, *9*(3), 239-261.
Doi: 10.1016/j.asw.2004.11.001

Şahin, M. D., Şen, S. and Güler, D. (2022). Adaptation and psychometric evaluation of the COVID-19 stress scales in Turkish sample. *International Journal of Assessment Tools in Education*, *9*(3), 593-612.
Doi: 10.21449/ijate.1067542

Şahin, M. G., Boztunç-Öztürk, N. and Taşdelen-Teker, G. (2015). Determining the pre-service teachers' measurement tool preferences for evaluation of their achievement. *Journal of Measurement and Evaluation in Education and Psychology*, *6*(1), 95-106.
Doi: 10.21031/epod.30481

Şahin-Sarkın, D. B. and Taşdemir, F. (2019). Scaling the information sought by the pre-service teachers in the test directive through method of sorting judgments. *Ankara University Journal of Faculty of Educational Sciences, 52*(2), 357-373.
Doi: 10.30964/auebfd.431803

Şen, S. and Göçen, A. (2021). A psychometric evaluation of the Ethical Leadership Scale using Rasch analysis and confirmatory factor analysis. *The Journal of General Psychology*, *148*(1), 84-104.
Doi: 10.1080/00221309.2020.1834346

Tat, O. and Anıl, D. (2016). A comparative analysis of Thurstone and Bradley-Terry scaling techniques. *Current Research in Education*, *2*(3), 174-185.

Toprak, E., Derin, S. and Güçlü, M. (2020). Characteristics to be possessed by a qualified student: Values or skills? *International Journal of Eurasian Education and Culture, 5*(8), 52-108.
https://www.ijoeec.com/DergiTamDetay.aspx?ID=85&Detay=Ozet

Torgerson, W. S. (1958). *Theory and methods of scaling*. Newyork: John Wiley & Sons.

Turgut, M. F. and Baykul, Y. (1992). *Scaling techniques [Ölçekleme teknikleri]*. Ankara: ÖSYM.

Turner, H. L., Van Etten, J., Firth, D. and Kosmidis, I. (2020). Modelling rankings in R: The plackettluce package. *Computational Statistics*, *35*, 1027-1057. Doi: 10.1007/s00180-020-00959-3

Uysal, M. (2015). *An investigation of psychometric properties of research self-efficacy scale according to classical test theory and item response theory* Unpublished master's thesis. Retrieved from YOK National Thesis Center. (Thesis ID: 388229)

Wind, S. and Hua, C. (2022). *Rasch measurement theory analysis in R*. Boca Raton, FL: CRC. Doi: 10.1201/9781003174660

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

Yalçın, S. and Şengül-Avşar, A. (2014). Scaling pedagogy courses faculty of education with rank-order judgments. *Journal of Measurement and Evaluation in Education and Psychology, 5*(2), 79-90. Doi: 10.21031/epod.71470

Yıldırım-Seheryeli, M. and Anıl, D. (2020). Scaling of mood-state and sample cases causing anger in a relationship with rank-order judgment and classifying judgment. *International Journal of Assessment Tools in Education*, *7*(1), 114-129. Doi: 10.21449/ijate.644805

Yücel, A. S. (2004). The analysis of the attitudes of secondary education students towards chemistry assignments. *Gazi University Journal of Gazi Education Faculty*, *24*(1), 147-159. https://dergipark.org.tr/en/download/article-file/77340

Zhu, W. and Yang, Y. (2016). Item response theory and its applications in Kinesiology. In N. Ntoumanis and N. D. Myers, (Eds.), *An introduction to intermediate and advanced statistical analyses for sport and exercise scientists* (57-77). Gloucester: Wiley.