International Journal of Progressive Education, Volume 19 Number 5, 2023

Comparison of Often Used Analysis Methods for Rank-Ordered Data

Süleyman Demir [i]
Sakarya University

Abstract

This study aims to compare the findings obtained from Rank-Ordered Judgments Scaling (ROJS), Placket-Luce Model (PLM), and Many Facet Rasch Model (MFRM) methods based on ranking judgments, which are often used in the analysis of rank-ordered data. For this purpose, one hundred senior students studying at the Faculty of Education and Faculty of Theology of Sakarya University were asked to rank pedagogical formation courses from the course they thought would be the most useful in their professional lives to the course they thought would be the least useful. The obtained data were analyzed using ROJS, PLM, and MFRM methods. When the obtained data were analyzed according to the ROJS, PLM, and MFRM, it was found that the course considered the least useful and the least preferred was the Instructional Technologies course. According to the raters, it was found that the most preferred and the most useful courses were Teaching Practice (I and II) in MFRM and ROJS, while in PLM, it was found to be the Classroom Management course. All other courses except the first-ranked course were sorted similarly in all models; the scale values in ROJS, logit values in MFRM, and worth in PLM were similar.

Keywords: Rank Ordered Data, Many Facet Rasch Model, Placket Luce Model, Judgment Scaling

DOI: 10.29329/ijpe.2023.603.1

Submitted: 13/03/2023 **Accepted:** 18/09/2023 **Published:** 16/10/2023

----

[i] **Süleyman Demir,** Assist. Prof. Dr., Department of Educatiaonal Measurement and Evaluation, Sakarya University, ORCID: 0000-0003-3136-0423

Email: suleymand@sakarya.edu.tr

INTRODUCTION

In social sciences and educational sciences, rank-ordered data are often used to determine the preferences or judgments of individuals. It is used in ordinal data as well as rank-ordered data in research. While ordinal data are the basis for sorting the respondents in the research group, rank-ordered data are the basis for sorting the options presented to the respondents. In rank-ordered data, respondents are mostly in the role of raters, judges, or referees. Therefore, in such studies, 'rater' is used instead of 'respondent.' In rank-ordered data, it may be asked to rank all n options (1^{st} to n^{th}) offered to raters, or it may be asked to rank the first three (1^{st} to 3^{rd}). The following have been determined using rank-ordered data; qualities that a teacher should have (Anıl & Güler, 2006), addictive substances (Kan, 2008), teacher competencies (Özer & Acar, 2011), factors affecting exam success (Bal, 2011), professional problems of teachers (Ekinci et al., 2012), teachers' assessment and evaluation method and tool preferences (Altun & Gelbal, 2014), pre-service teachers' social activity preferences (Polat & Göksel, 2014), pre-service teachers' internet usage preferences (Albayrak Sarı & Gelbal, 2015), patients' treatment preferences (Shakir et al., 2021), culturally sensitive teacher characteristics (Sarıdaş & Nayir, 2021), reasons for choosing university (Koçak & Çokluk-Bökeoğlu, 2021), urban mobility problems and transportation solutions (Kijewska et al., 2022). The studies conducted show that there are quite a lot of areas of use for rank-ordered data.

It is observed that ordered data are often analyzed based on ranking judgments, a scaling method. Scaling is defined as determining the psychological responses of stimuli in the physical dimension (Baykul and Turgut, 1992). On the other hand, rank-ordered judgment scaling (ROJS) allows the stimuli presented to the raters to be displayed on a scale based on the evaluator's judgments and the distances between the stimuli to be determined. For the analysis of rank-ordered data based on ranking judgments, the following steps can be followed:

- A sequence frequency table is created for the judgments of the raters.
- A ratios matrix is obtained using each sequence frequency in the sequence frequency table.
- The ratio in each cell in the ratio's matrix is converted into a Unit Normal Deviations Matrix using the unit normal distribution function.
- The scale values for each stimulus are obtained by averaging the values in each column of the Unit Normal Deviations Matrix.
- All scale values are shifted to standardize the scale values so that the smallest scale value is zero (Anil & Inal, 2017; Baykul & Turgut, 1992).

Since there is no program or open code for analyzing rank-ordered data based on ranking judgments, researchers must perform statistical calculations manually or using the Microsoft Excel calculation tool.

Another method used to analyze rank-ordered data is the Placket-Luce Model (PLM), which was developed by Plackett (1975) based on Luce's axiom (Luce, 1959) (based on the probability of choosing n an item among the number of stimuli i^{th}). In PLM, the probabilities of ranking the stimuli in a particular order are used to estimate the worth of stimuli related to each stimulus and express the importance of the stimulus. The fact that the value is high indicates that this stimulant is more important than other stimulants. Maximum probability or Bayesian methods are used for estimating the worth of stimuli (Turner et al., 2021). The value obtained from the analysis is divided by the standard error, and standard z scores are calculated for each stimulus. According to the statistically significant standard z scores, information is obtained that stimulants differ from the average (Finch, 2022; Turner et al., 2021). The analyses related to PLM can be performed in R using the PlacketLuce function developed by Turner et al. (2021) (Finch, 2022).

According to both ROJS and PLM, the raters' judgments are used to analyze the rank-ordered data. Another method used to analyze the evaluator's judgments is the Many Facet Rasch Model (MFRM). According to Linacre (1994), when analyzing the MFRM, all sources of variance (individual abilities, raters, criteria, etc.) being included in the analysis, it is possible to reach measurement results that are free from these sources of variance (Linacre, 1994). In other words, the measurement results of the students with MFRM can be determined independently of the raters and the criteria (or questions) (Farrokhi & Esfandiari, 2011; Linacre, 1994). With MFRM, it is possible to analyze multi-category (polytomous) data obtained using Likert Scale, Semantic Differential Scale, and Open-ended items (Ilhan, 2016; Knoch & McNamara, 2015; Linacre, 1994), as well as rank-ordered data (Linacre, 1994) and paired comparison data (Linacre, 1994; Linacre, 2006). MFRM analyses can be performed with the Facets program developed by Linacre (2014) and the R package that Robitzsch et al. (2022) prepared.

In order for the findings obtained as a result of the analyses to be more accurate, the most important point is to select the analysis method to be used correctly. In case there is more than one method to perform the analysis, it is necessary to choose the one that gives the most accurate results and is the most economical and practical from the user's point of view. When the literature is examined, it is seen that although the PLM and ROJS models are used only in rank-ordered data (Baykul & Turgut, 1992; Guilford, 1954; Finch, 2022; Turner et al., 2021), the MFRM model can also be used in dichotomous and Likert type data (Linacre, 1994; Linacre, 2006). Although each method can be used on rank-ordered data, no findings have been found in the literature regarding the differences in the mathematical infrastructures of these models, how they will affect the analysis results, and the differences or similarities between the analysis results. This study aims to compare the findings obtained from the ROJS, PLM, and MFRM methods, which are often used in the analysis of rank-ordered data.

METHOD

Participants and the data

To achieve the purpose of the study, it was asked to rank the pedagogical formation courses that the students of the faculty of education and the faculty of theology took during their undergraduate education. The research data consists of one hundred senior students who are studying at the Faculty of Education and Faculty of Theology of Sakarya University during the 2022-2023 academic term. In determining the students who would be included in the research group, the criterion of having taken courses other than the Guidance and Special Education courses (Guidance and Special Education courses are not included in the eighth semester of undergraduate class plans, students have not taken these courses yet or they are not included because they are taking the data collection process). Eight pedagogical formation courses given jointly to the students at the faculty of education and the faculty of theology were presented (Introduction to Education, Educational Psychology, Teaching Principles and Methods, Classroom Management, Measurement and Evaluation, Teaching Practices (I and II), Special Teaching Methods, Instructional Technologies). Before applying the data collection tool, ethics committee approval was obtained with the decision of Sakarya University Educational Research and Publication Ethics Committee dated 02.15.2023 and numbered E-61923333-050.99-222305. Students have ranked the pedagogical formation courses so that “1” is for the course they think will be the most useful in their professional life, and “8” is for the course they think will be the least useful.

Data Analysis

In this section, the findings related to ROJS, PLM, and MFRM, which are used in the analysis of the data, are given in the following sections.

Rank-ordered judgment scaling

The ROJS model is a method first developed by Guilford (1954). According to this model, the scale values of stimulants are obtained based on the data obtained as rank ordered. In order to obtain the scale values, first of all, the Ranking Matrix (1) is created, which contains the ranking made by the m-score raters for n stimuli. The rankings matrixshows the ranking result of l^{th} raters and x j^{th} stimuli.

$$\begin{array}{c} \\ J_1 \\ J_2 \\ J_l \\ J_m \end{array}\begin{array}{cccc} S_1 & S_2 & S_j & S_n \\ \left[\begin{array}{cccc} R_{11} & R_{12} & \dots & R_{1n} \\ R_{21} & R_{22} & \dots & R_{2n} \\ \dots & \dots & R_{ij} & \dots \\ R_{m1} & R_{m2} & \dots & R_{mn} \end{array}\right. \end{array} \quad (1)$$

Using the ordering matrix, the Sequence Frequencies Matrix (2) is created, which contains the sequence frequencies of the stimuli. The Sequence Frequencies Matrix, F_{ik}, k^{th} represents the frequency of the stimulus being displayed in i^{th} order by the raters.

$$\begin{array}{c|cccc} & S_1 & S_j & S_k & S_n \\ \hline 1 & F_{11} & F_{1j} & F_{1k} & F_{1n} \\ 2 & F_{21} & F_{2j} & F_{2k} & F_{2n} \\ i & F_{i1} & F_{ij} & F_{ik} & F_{in} \\ n & F_{n1} & F_{nj} & F_{nk} & F_{nn} \end{array} \quad (2)$$

After this stage, the rank-ordered data can be scaled by converting it into a scaling method based on binary comparisons. This transformation is calculated as how much higher the order of a stimulus is than the order of other stimuli. In other words, for two different stimuli, such as S_j and S_k, S_j stimulus is preferred more than S_k stimulus in i^{th} order, $\left(n\left(S_{ij} > S_{ik}\right)\right)$ is calculated as follows (3):

$$n\left(S_{ij} > S_{ik}\right) = F_{ij}.\left(F_{k<i} + \left(1/2\right)F_{ik}\right) \quad (3)$$

n The probability that the S_j stimulus is preferred over the S_k stimulus is obtained (4) by summing the calculated $n\left(S_{ij} > S_{ik}\right)$ for each row and dividing $\left(m^2\right)$ by the square of the rater number.

$$\begin{bmatrix} & S_1 & S_2 & \dots & S_j & \dots & S_n \\ S_1 & & P_{12} & & P_{1j} & & P_{1n} \\ S_2 & P_{21} & & & P_{2j} & & P_{2n} \\ \vdots & & & & & & \\ S_j & P_{j1} & P_{j2} & & & & P_{jn} \\ \vdots & & & & & & \\ S_n & P_{n1} & P_{n2} & & P_{nj} & & \end{bmatrix} \tag{4}$$

The $\left(z_{jk}\right)$ values corresponding to the areas under each ratio $\left(P_{jk}\right)$ unit normal distribution function in the ratios matrix are calculated using the conversion tables (Edwards, 1957, pp: 246-247). Scale values are obtained by averaging the $n-1$ z values of each stimulus (Anıl & İnal, 2017; Baykul & Turgut, 1992; Edwards, 1957; Guilford, 1954). It shows that as the scale values of stimulants increase, stimulants are given a relatively higher rank value to other stimulants. In other words, it is relatively less preferred to other stimulants.

Placket-Luce Model

The Placket-Luce model (PLM) was developed using the Luce axiom (Luce, 1959). The Luce axiom is based on the probability of choosing i^{th} item among n stimuli set (S). This possibility is as follows;

$$P(i:S) = \frac{a_i}{\sum_{i \in S} ai} \tag{5}$$

When K raters are asked to rank *n* stimuli, the rater must first choose the first stimulus among n stimuli and then choose the second stimulus among the remaining *n-1* stimuli. This process continues until the selection of the last stimulus. The probability of ranking the stimuli in a certain order (*w*) (where $A_i = i$ is the vector of alternative rankings which is selected in j^{th} rank) is defined as follows:

$$P(w) = \prod_{i=1}^{n} \frac{a_{ij}}{\sum_{i \in A_j} ai} \tag{6}$$

The estimation of the worth of stimuli for each stimulus based on this probability can be made by Maximum Likelihood or Bayesian Estimation methods. PLM analyses can be performed in R with the PlackettLuce function written by Turner (2022). With the PlackettLuce function, the worth of stimuli is calculated in two ways. In the first method, one of the stimulants is taken as a reference. The value of the stimulus is fixed to zero for the stimulus taken as a reference. For stimulants other than reference, the worth of stimuli is estimated depending on the reference. The second method is also taken as a reference to the mean of values. For all stimulants, worth is estimated as a mean of values reference. Since the value estimated by both methods is calculated based on reference, the interpretation of stimuli is relative (Finch, 2022).

In addition to the worth of stimuli, the standard error is also calculated for each worth. These standard errors are used to evaluate the significance of the worth of stimuli. If the value is estimated by referring to the first stimulus, it is determined whether the other stimuli differ significantly from the referenced stimulus. If the worth is estimated by reference to the mean of values, it is determined whether all stimulators have a significant difference according to the mean of values (Finch, 2022; Turner, 2022).

Many Facet Rasch Models

MFRM is an extended version of the one-parameter Item Response Theory model, also known as the Rasch model. The model (7) developed by Rasch (1980) is used for two categories of substances and consists of the facets of substance difficulty and individual ability.

$$\log\left(P_{ni1}/P_{ni0}\right) = B_n - D_i \tag{7}$$

The model developed by Rasch was later adapted by Andrich (1978) for Likert-type substances and by Masters (1982) for partial credit items (Linacre, 1994). Linacre (1994) developed the MFRM by adding variance sources likely to affect the measurement results, such as raters' judgments, to the model adapted for partial credit items (İlhan, 2016). Different models can be created according to the variance sources added in MFRM. An example of a three-facet model is given below (8).

$$\log\left(P_{nijk}/P_{nij(k-1)}\right) = B_n - D_i - C_j - F_k \tag{8}$$

In this model;

P_{nijk} : Probability of j^{th} raters giving k value to i^{th} item of n^{th} individual

$P_{nij(k-1)}$: The probability of j^{th} raters giving the k-1 value to i^{th} item of n^{th} individuals

B_n : the ability of n^{th} individual

D_i : i^{th} difficulty of the item

C_j : the firmness/generosity of the j^{th} raters

F_k : the difficulty of the step up from $(k-1)^{th}$ category to k^{th} category (İlhan, 2016; Linacre, 1994).

Linacre (1994) has developed a formula (9) that can be used for MFRM for rank-ordered data based on the seventh formula. The function of the ranking of n stimulants of K rater in a $(S_1, S_2, ..., S_n)$ certain order is determined by $\left(\log\left(R\left(n\right)\right)\right)$

$$\log\left(R\left(n\right)\right) = \log\left(\prod_{r=1}^{K} \frac{\prod_{j=1}^{n}\prod_{k=j+1}^{n} X_{rjk} * \exp\left(S_j\right) + X_{rkj} * \exp\left(S_k\right)}{\sum_{s=1}^{n!}\prod_{j=1}^{n}\coprod_{k=j+1}^{n} X_{sjk} * \exp\left(S_j\right) + X_{rskj} * \exp\left(S_k\right)}\right) \tag{9}$$

X_{rjk} : It is equal to 1 if the k rater gives a higher rank value to the stimulus $\left(S_j\right)$ than the stimulus $\left(S_k\right)$, and 0 if it gives a smaller rank value.

X_{rkj} : It is equal to 1 if the k rater gives a higher rank value to the stimulus (S_k) than the stimulus (S_j), and 0 if it gives a smaller rank value.

FINDINGS

The findings of the Rank-Ordered Judgement Scaling Model

In order to be able to analyze the data with the ROJS Model, the calculated frequency statistics regarding the order in which each stimulant is preferred are given in Table 1.

Table 1: Frequency matrix of the stimuli

Rank	A	B	C	D	E	F	G	H
1	13	8	4	13	4	49	5	4
2	7	22	14	29	6	7	12	3
3	6	10	22	22	12	7	12	9
4	9	11	15	12	20	5	17	11
5	13	7	19	12	20	4	13	12
6	11	17	12	6	11	14	13	16
7	9	16	9	4	13	7	23	19
8	32	9	5	2	14	7	5	26

A=Introduction to Education, B=Educational Psychology, C=Teaching Principles and Methods, D=Classroom Management, E=Measurement and Evaluation, F=Teaching Practices (I and II), G=Special Teaching Methods, H=Instructional Technologies.

When Table 1 is examined, it is seen that 13 students prefer the Introduction to Education course, 8 students who prefer the Educational Psychology course, 4 students who prefer Teaching Principles and Methods course, 13 students who prefer the Classroom Management course, 4 students who prefer Measurement and Evaluation course, 49 students who prefer Teaching Practices (I and II) course, 5 students who prefer Special Teaching Methods course, 4 students who prefer Instructional Technologies course. The Ratios Matrix in Table 2 is obtained by determining how much the ranking of each stimulus is greater than the ranking of other stimuli using formula 3.

Table 2: The Ratios Matrix

	A	B	C	D	E	F	G	H
A		0.395	0.364	0.268	0.443	0.270	0.414	0.536
B	0.605		0.479	0.350	0.570	0.320	0.540	0.667
C	0.636	0.521		0.336	0.588	0.348	0.512	0.729
D	0.732	0.651	0.664		0.750	0.420	0.714	0.819
E	0.557	0.430	0.412	0.250		0.278	0.468	0.616
F	0.731	0.680	0.652	0.580	0.722		0.706	0.783
G	0.586	0.460	0.488	0.286	0.532	0.294		0.646
H	0.464	0.333	0.271	0.181	0.384	0.218	0.354	

A=Introduction to Education, B=Educational Psychology, C=Teaching Principles and Methods, D=Classroom Management, E=Measurement and Evaluation, F=Teaching Practices (I and II), G=Special Teaching Methods, H=Instructional Technologies.

When Table 2 is examined, the value in each cell shows the probability that the stimulator in the row will be preferred over the stimulator in the column. While the probability that the Introduction to Education course will be preferred over the Educational Psychology course is 0.395, the probability that the Educational Psychology course will be preferred over the Introduction to Education course is 0.605. The unit normal deviations matrix in Table 3 is obtained using each ratio in the ratios matrix. By taking the average of the values in each row, the average values for each stimulus are calculated. Scale values are obtained by moving the smallest of the average values so that it is zero.

Table 3: Unit Normal Deviations Matrix and Scale values

	A	B	C	D	E	F	G	H
A		-0.266	-0.347	-0.620	-0.143	-0.614	-0.217	0.091
B	0.266		-0.052	-0.387	0.176	-0.468	0.101	0.431
C	0.347	0.052		-0.422	0.222	-0.390	0.030	0.610
D	0.620	0.387	0.422		0.675	-0.202	0.565	0.913
E	0.143	-0.176	-0.222	-0.675		-0.589	-0.081	0.294
F	0.614	0.468	0.390	0.202	0.589		0.541	0.781
G	0.217	-0.101	-0.030	-0.565	0.081	-0.541		0.375
H	-0.091	-0.431	-0.610	-0.913	-0.294	-0.781	-0.375	
Mean	0.265	-0.008	-0.056	-0.423	0.163	-0.448	0.070	0.437
Scale Value	0.713	0.440	0.392	0.026	0.611	0.000	0.519	0.885

A=Introduction to Education, B=Educational Psychology, C=Teaching Principles and Methods, D=Classroom Management, E=Measurement and Evaluation, F=Teaching Practices (I and II), G=Special Teaching Methods, H=Instructional Technologies.

When Table 3 is examined, it is seen that the lowest scale value belongs to the Teaching Practices (I and II) course, while the highest scale value belongs to the Instructional Technologies course. In other words, it seems that the course that students think will be most useful for their professional lives is the Teaching Practices (I and II) course, while the course they think will be least useful is the Instructional Technologies course.

The Findings of the Plackett Luce Model

In order to analyze the data with PLM, the PlackettLuce package (Turner, 2022) was used. When the mean of values is taken as a reference in Table 4, the calculated values of worth, standard error, z-value, and significance are given for each stimulus.

Table 4: Value, standard error, z-value, and significance values calculated when referencing the mean of values (PLM_M)

	Worth	Std..Error	z	p
Introduction to Education (A)	-0.578	0.119	-4.835	0.000
Educational Psychology (B)	-0.037	0.110	-0.339	0.734
Teaching Principles and Methods (C)	0.192	0.107	1.796	0.073
Classroom Management (D)	0.725	0.109	6.636	0.000
Measurement and Evaluation (E)	-0.198	0.109	-1.818	0.069
Teaching Practices (I and II) (F)	0.587	0.115	5.105	0.000
Special Teaching Methods (G)	-0.084	0.107	-0.784	0.433
Instructional Technologies (H)	-0.608	0.114	-5.307	0.000

According to Table 4, the lowest worth belongs to the Instructional Technologies course, and the highest belongs to the Classroom Management course. It is seen that the course that students think will be the least useful for their professional lives is the Instructional Technologies course, while the course that they think will be the most useful is the Classroom Management course. In addition, it is observed that Introduction to Education and Instructional Technologies courses are statistically significantly lower than the mean of values, and Classroom Management and Teaching Practices (I and II) courses are statistically significantly higher than the mean of values. Educational Psychology, Teaching Principles, Methods, Measurement, Evaluation, and Special Teaching Methods courses do not differ statistically significantly from the mean values. Table 5 gives the values of worth, standard error, z-value, and significance calculated for each stimulus when the first stimulus is taken as a reference (It was not calculated for this system because the Introduction to Education course was taken as a reference).

Table 5: Value, standard error, z-value, and significance values calculated when the first stimulus is referenced (PLM_1)

	Worth	Std..Error	z	p
Introduction to Education (A)	NA	NA	NA	NA
Educational Psychology (B)	0.540093	0.168824	3.199146	0.001378
Teaching Principles and Methods (C)	0.768925	0.171223	4.490785	0.000007
Classroom Management (D)	1.302245	0.17619	7.39112	0.000000
Measurement and Evaluation (E)	0.378987	0.172769	2.193612	0.028263
Teaching Practices (I and II) (F)	1.164565	0.182648	6.375995	0.000000
Special Teaching Methods (G)	0.49366	0.172498	2.861834	0.004212
Instructional Technologies (H)	-0.03025	0.177226	-0.1707	0.864463

According to Table 5, the lowest worth belongs to the Instructional Technologies course, and the highest belongs to the Classroom Management course. Similar to the previous findings, it is seen that the course that students think will be the least useful for their professional lives is the Instructional Technologies course. In contrast, the course that they think will be the most useful is the Classroom Management course. It is observed that the worth of Educational Psychology, Teaching Principles and Methods, Classroom Management, Measurement and Evaluation, Teaching Practices (I and II), and Special Teaching Methods courses is statistically significantly higher than the worth of the Introduction to Education course. The worth of the Instructional Technologies course does not differ statistically significantly from the worth of the Introduction to Education course.

The findings of Many Facet Rasch Model

Before the data analysis with MFRM, the ratio of unexpected results or extreme values was examined in order to assess the model data fit. The ratio of unexpected results or extreme values outside the range of ±2 and ±3 by Linacre (2014) should be less than 5% and 1%, respectively. When the analysis results are examined, the rate of those outside the range of ±2 is 3.25%, and those outside the range of ±3 is 0.00%. Accordingly, model data fit is provided for analysis with MFRM. Figure 1 shows the data calibration map obtained as a result of analyzing the judgments related to the ranking of eight courses by 100 raters.

```
+---------------------------------+
|Measr|+raters     |-stimuli |Scale|
|-----+------------+---------+-----|
|   1 +            +         + (8) |
|     |            |         |     |
|     |            |         |     |
|     |            |         |  7  |
|     |            |         |     |
|     |            |         | --- |
|     |            |         |     |
|     |            | D  F    |  6  |
|     |            |         | --- |
|     |            |         |  5  |
*   0 * ********** * B  C    * --- *
|     |            | E  G    |  4  |
|     |            | A       | --- |
|     |            | H       |  3  |
|     |            |         |     |
|     |            |         | --- |
|     |            |         |     |
|     |            |         |  2  |
|     |            |         |     |
|     |            |         |     |
|  -1 +            +         + (1) |
|-----+------------+---------+-----|
|Measr| * = 10     |-stimuli |Scale|
+---------------------------------+
```

Figure 1: Data calibration map[1]

In the data calibration map, the columns contain information about the facets. As you go down from top to bottom in the rater column, the strictness of the raters increases. The stimuli column shows that stimulants' preference order increases as they go down from top to bottom. In other words, the

[1] A=Introduction to Education, B=Educational Psychology, C=Teaching Principles and Methods, D=Classroom Management, E=Measurement and Evaluation, F=Teaching Practices (I and II), G=Special Teaching Methods, H=Instructional Technologies.

preference rates decrease. The data calibration map shows that all raters have the same rigidity/generosity on the rater facet (All raters rated from 1 to 8 when rating the stimuli). On the stimuli facet, it shows that stimulus-D (Classroom Management) and stimulus-F (Teaching Practices (I and II)) are most preferred compared to other stimuli, and stimulus-H (Instructional Technologies) is preferred the least compared to other stimuli.

In Figure 2, the analysis results regarding the ranking of the stimuli by the raters are shown. The logit value of stimulus-F is 0.30; the logit value of the stimulus-D is 0.27, and the logit value of the stimulus-H is -0.28.

```
+---------------------------------------------------------------------------------------------------------------------+
| Total   Total   Obsvd  Fair-M|        Model | Infit      Outfit    |Estim.| Correlation |                           |
| Score   Count  Average Avrage|Measure  S.E. | MnSq ZStd  MnSq ZStd|Discrm| PtMea PtExp | N stimulus                |
|------------------------------+--------------+----------------------+------+-------------+---------------------------|
|  313    100      3.13   3.13|   .30    .05 | 1.65  4.2  1.65  4.2|  .86 |   .00   .00 | F                         |
|  325    100      3.25   3.25|   .27    .05 |  .75 -2.1   .75 -2.1| 1.06 |   .00   .00 | D                         |
|  428    100      4.28   4.28|   .04    .05 |  .68 -3.8   .68 -3.8| 2.83 |   .00   .00 | C                         |
|  447    100      4.47   4.47|   .00    .05 | 1.04   .4  1.04   .4|  .39 |   .00   .00 | B                         |
|  477    100      4.77   4.77|  -.06    .05 |  .82 -2.0   .82 -2.0| 1.78 |   .00   .00 | G                         |
|  501    100      5.01   5.01|  -.11    .05 |  .79 -2.3   .79 -2.3| 1.33 |   .00   .00 | E                         |
|  531    100      5.31   5.31|  -.17    .05 | 1.39  3.3  1.39  3.3|  .76 |   .00   .00 | A                         |
|  578    100      5.78   5.78|  -.28    .05 | 1.01   .1  1.01   .1| 1.00 |   .00   .00 | H                         |
|------------------------------+--------------+----------------------+------+-------------+---------------------------|
|  450.0  100.0    4.50   4.50|   .00    .05 | 1.02  -.3  1.02  -.3|      |   .00       | Mean (Count: 8)           |
|   87.4     .0     .87    .88|   .19    .00 |  .32  2.7   .32  2.7|      |   .00       | S.D. (Population)         |
|   93.5     .0     .93    .94|   .20    .00 |  .34  2.9   .34  2.9|      |   .00       | S.D. (Sample)             |
+---------------------------------------------------------------------------------------------------------------------+
Model, Populn: RMSE .05  Adj (True) S.D. .18  Separation 3.88  Strata 5.50  Reliability .94
Model, Sample: RMSE .05  Adj (True) S.D. .20  Separation 4.16  Strata 5.88  Reliability .95
Model, Fixed (all same) chi-square: 117.0  d.f.: 7  significance (probability): .00
Model,  Random (normal) chi-square: 6.6  d.f.: 6  significance (probability): .36
-----------------------------------------------------------------------------------------------------------------------
```

A=Introduction to Education, B=Educational Psychology, C=Teaching Principles and Methods, D=Classroom Management, E=Measurement and Evaluation, F=Teaching Practices (I and II), G=Special Teaching Methods, H=Instructional Technologies.

Figure 2: The results of the analysis of the order of stimulants by the raters

When Figure 2 is examined, stimulus-F (Teaching Practices (I and II)) and stimulus-D (Classroom Management) were preferred the most, and stimulus-H (Instructional Technologies) and stimulus-A (Introduction to Education) were preferred the least. The separation ratio calculated for the stimulus facet, the reliability coefficients between 3.88 and 0.94, and the chi-square statistics being significant (X^2=117.00 sd=7, p<0.05) show that the stimuli are differentiated from each other at a statistically significant level.

Table 6: Rankings obtained according to ROJS, PLM, and MFRM

	ROJS		MFRM		PLM_M		PLM_1	
1.	F	0.000	F	0.300	D	0.725	A	NA
2.	D	0.026	D	0.270	F	0.587	D	1.302
3.	C	0.392	C	0.040	C	0.192	F	1.165
4.	B	0.440	B	0.000	B	-0.037	C	0.769
5.	G	0.519	G	-0.060	G	-0.084	B	0.540
6.	E	0.611	E	-0.110	E	-0.198	G	0.494
7.	A	0.713	A	-0.170	A	-0.577	E	0.379
8.	H	0.885	H	-0.280	H	-0.608	H	-0.030

A=Introduction to Education, B=Educational Psychology, C=Teaching Principles and Methods, D=Classroom Management, E=Measurement and Evaluation, F=Teaching Practices (I and II), G=Special Teaching Methods, H=Instructional Technologies, NA= No Answer.

When TTable 6 is examined, in ROJS and MFRM models, stimulus-F (Teaching Practices (I and II)) was preferred the most, and stimulus-H (Instructional Technologies) was preferred the least. In PLM models, stimulus-D (Classroom Management) was preferred the most, and stimulus-H (Instructional Technologies) was preferred the least.

DISCUSSION AND RESULTS

This research aims to compare the findings obtained from the ROJS, PLM, and MFRM methods, which are often used in analyzing rank-ordered data. For this purpose, the raters were asked to rank the pedagogical formation courses they took during their undergraduate education in the form of the course that they thought would be the most useful in their professional lives and the course that would be the least useful. When the obtained data were analyzed according to the ROJS, PLM, and MFRM, it was found that the course considered the least useful and the least preferred course was the Instructional Technologies course. In future studies, studies can be conducted to investigate why students prefer the "classroom management" course or the "teaching practice" course.

It was found that the most preferred course that was considered by the raters was Teaching Practice (I and II) courses in MFRM and ROJS, while Classroom Management courses were found to be in PLM. As can be seen in Table 1, the reason for this difference is that although Teaching Practice (I and II) courses were preferred by many raters in the first place, they were not preferred in the 2nd, 3rd, 4th, and 5th ranks. Although the raters who prefer the Classroom Management course in the first place are relatively few, the number of raters who prefer it in the 2nd, 3rd, 4th, and 5th places is high. This may be why the classroom management course is in the first place while the teaching practice course is in the second place in PLM. All other courses except the first-ranked course were sorted in the same way in all models, and the scale values in ROJS, logit values in MFRM, and worth in PLM were obtained similarly. In this case, it can be said that the analysis results of all three models are similar.

Like the findings obtained in this study, the logit values calculated in the Rasch analysis and scale values calculated in the ROJS analysis made by Wainer et al. (1978) were close to each other. In addition, a study conducted by Güler et al. (2018) found that the scale values obtained by the scaling method based on binary comparisons and the logit values obtained from MFRM were similar. Similar studies in the literature also support the finding obtained from this study that different methods yield similar results. It can be said that the reason why similar results were obtained in the MFRM and ROJS models in this study and the existing studies in the literature is that both models perform analyses in a single stage, and the order in which a stimulus is selected is included in the analysis with equal probability.

Based on the research findings, comparable results will be obtained when all three models are used to analyze bank-ordered data. Although all three models give comparable results, there are some advantages and disadvantages. Although using the Microsoft Office Excel program to analyze the data with the ROJS method makes it easier for researchers to access the program, the formulas must be rewritten each time to analyze the data. In order to analyze the data with MFRM, it is performed with the FACET program developed by Linacre (2014) or the R package prepared by Robitzsch et al. (2022), and the analyses are performed via syntax. In order to analyze the data with PLM, it is performed with the PlacketLuce R package developed by Turner et al. (2021). While the significance of the differences between the stimuli cannot be evaluated in ROJS and MFRM, the significance of the stimuli according to the mean and the 1st stimulus can be assessed in PLM. In addition, in PLM, it can also be determined whether there is a difference between one stimulus and another. For this purpose, one of the two stimulants to be evaluated should be determined as a reference and analyzed. Researchers should consider the advantages and disadvantages of the models when using these models.

Conflicts of Interest: No potential conflict of interest was declared by the authors.

Funding Details: This study was not supported by any funding agency

CRediT Author Statement: The author confirms sole responsibility for the following: study conception and design, data collection, analysis and interpretation of results, and manuscript preparation.

Ethical Statement: Ethical permission was taken from the Sakarya University Educational Research and Publication Ethics Committee. 15.02.2023 and numbered E-61923333-050.99-222305

REFERENCES

Albayrak-Sarı, A. & Gelbal, S. (2015). İkili karşılaştırmalar yargılarına ve sıralama yargılarına dayalı ölçekleme yaklaşımlarının karşılaştırılması [A comparison of scaling procedures based on pair-wise comparison and rank-order judgments scaling]. *Journal of Measurement and Evaluation in Education and Psychology, 6*(1), 126-141. http://dx.doi.org/10.21031/epod.30288.

Altun, A. & Gelbal S. (2014). Öğretmenlerinin kullandıkları ölçme ve değerlendirme yöntem veya araçlarının ikili karşılaştırma yöntemiyle belirlenmesi [Determining teachers' measurement tools or techniques via pair-wise comparison method]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, *2*(2), 200-209. https://doi.org/10.21031/epod.09064.

Anıl, D. & Güler, N. (2006). İkili karşılaştırma yöntemi ile ölçekleme çalışmasına bir örnek [An example of the scaling study by pair-wise comparison method]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 30,* 30-36. http://efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/722-published.pdf.

Bal, Ö. (2011). Seviye Belirleme Sınavı (SBS) başarısında etkili olduğu düşünülen faktörlerin sıralama yargıları kanunuyla ölçeklenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 2*(2), 200-209. https://dergipark.org.tr/tr/download/article-file/65975.

Ekinci, A., Bindak, R. & Yıldırım, C. (2012). İlköğretim okulu yöneticilerinin öğretmenlerin mesleki sorunlarına empatik yaklaşımlarının ikili karşılaştırmalar metodu ile incelenmesi [A research regarding the empathic approaches of school managers about professional problems of teachers by pairwise comparisons method]. *Gaziantep Üniversitesi Sosyal Bilimler Dergisi. 11*(3), 759-776. https://dergipark.org.tr/tr/download/article-file/223316.

Farrokhi, F., & Esfandiari, R. (2011). A many-facet Rasch model to detect halo effect in three types of raters. *Theory and Practice in Language Studies, 1*(11), 1531–1540. http://dx.doi.org/10.4304/tpls.1.11.1531-1540

Finch, H.. (2022). An introduction to the analysis of ranked response data. *Practical Assessment, Research & Evaluation, 27*(7), 1-20. https://scholarworks.umass.edu/pare/vol27/iss1/7/

Güler, N., İlhan, M. & Taşdelen-Teker, G. (2018). İkili karşılaştırmalarla ölçekleme yöntemi ile Rasch analizinden elde edilen ölçek değerlerinin karşılaştırılması [Comparing the scale values obtained from pairwise comparison scaling method and rasch analysis]. *İnönü Üniversitesi Eğitim Fakültesi Dergisi, 19*(1), 31-48. http://dx.doi.org/10.17679/inuefd.400386.

İlhan, M. (2016). A comparison of the results of many-facet Rasch analyses based on crossed and judge pair designs. *Educational Sciences: Theory & Practice, 16*(2), 579-601. http://dx.doi.org/10.12738/estp.2016.2.0390.

Kan, A. (2008). Yargıcı kararlarına dayalı ölçekleme yöntemlerinin karşılaştırılması üzerine ampirik bir çalışma [A comparision of scaling methods based on judge decisions: an empirical study]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 35,* 186-194. http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/552-published.pdf.

Kijewska, K., França, J. G. C. B., de Oliveira, L. K., & Iwan, S. (2022). Evaluation of urban mobility problems and freight solutions from residents' perspectives: a comparison of Belo Horizonte (Brazil) and Szczecin (Poland). *Energies, 15*(3), 710. https://doi.org/10.3390/en15030710.

Knoch, U., &, McNamara, T. (2015). Rasch analysis. In L. Plonsky, (Ed.), Advancing quantitative methods in second language research (pp. 275–304). Routledge.

Koçak, D. & Çokluk-Bökeoglu, Ö. (2021). Üniversite Tercih Nedenlerinin İkili Karşılaştırma ve Sıralama Yargıları Yöntemleri ile Ölçeklenmesi [Scaling university preferences with pair-wise comparison method and rank-order judgments method]. *MANAS Sosyal Araştırmalar Dergisi, 10*(3), 1580-1591 . DOI: https://doi.org/10.33206/mjss.874504

Linacre, J. M. (1994). Many-Facet Rasch Measurement. Chicago: Mesa Press.

Linacre J. M. (2006). Rasch Analysis of Rank-Ordered Data. *Journal of Applied Measurement*, *7*(1), 129-136. https://pubmed.ncbi.nlm.nih.gov/16385155/.

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

Özer, Y. & Acar, M. (2011). Öğretmenlik mesleği genel yeterlikleri üzerine ikili karşılaştırma yöntemiyle bir ölçekleme çalışması. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi, 3*(40), 89-101.

Robitzsch, A., Kiefer, T., & Wu, M. (2022). TAM: Test Analysis Modules. R package version 4.1-4, https://CRAN.R-project.org/package=TAM.

Sarıdaş, G. & Nayir, F. (2021). Kültürel değerlere duyarlı öğretmen özelliklerinin sıralama yargılarıyla ölçeklenmesi [Scaling Culturally responsive teacher characteristics with ranking judgments]. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 53,* 355- 377. https://doi.org/10.9779.pauefd.827009.

Shakir, M. K. M., Brooks, D. I., McAninch, E. A., Fonseca, T.L., Mai, V. Q., Bianco, A. C. & Thanh, D. H. (2021). Comparative effectiveness of levothyroxine, desiccated thyroid extract, and levothyroxine+liothyronine in hypothyroidism. *The Journal of Clinical Endocrinology & Metabolism, 106*(11). https://doi.org/10.1210/clinem/dgab478.

Turner, H., Kosmidis, I., Firth, D., & van Etten, J. (2021). Modelling rankings in R: The Plackettluce package. *Computational Statistics*, *35,* 1027-1057. https://doi.org/10.1007/s00180-020-00959-3.