Using Rank-order Judgments Scaling to Determine Students' Evaluation Preferences*

Durmus OZBASI[1]

ARTICLE INFO	ABSTRACT
	Purpose: This study sought to determine university students' evaluation preferences and then scaled them based on their rank-order judgments. **Research Methods**: The survey model was used in this study. This study was conducted with a total of 376 university students of varying grade levels enrolled in different departments of the faculty of education of two separate state universities in Turkey during the 2017-2018 academic year. Data were collected using a 13-item survey designed specifically for this study that solicited answers regarding students' evaluation preferences in measuring their academic performance. Students first ranked to evaluation types from most to least preferred and then assigned a single number for each stimulus. The data attained from the study were then scaled based on rank-order judgments.

Findings: The study findings revealed that students most preferred to be assessed using oral exams and least preferred tests composed of multiple-choice questions. **Implications for Research and Practice**: This study was restricted to university students enrolled in the faculty of education of two state universities in Turkey. By conducting a similar study with students enrolled in other faculties in the same or different higher education institutions results and potential differences between faculties may be compared.

---

[1] Corresponding Author, Çanakkale Onsekiz Mart University, TURKEY, e-mail: dozbasi@comu.edu.tr ,
ORCID ID: https://orcid.org/0000-0001-5078-477X

Introduction

Current technology has developed at such remarkable speeds that access to information has been made profoundly easy, and this has forced educational systems to adjust themselves to fit the current age. Educational systems change in line with this and develop appropriate strategies conforming to prevailing pedagogical methods and techniques while also seeking to cultivate in students the type of human strengths needed in the current age. Consequently, education systems play a critical role within the greater education process in raising qualified individuals equipped with the knowledge and skills necessary to succeed in life.

One of the fundamental objectives of the educational process is, undoubtedly, to instill a set of desired behaviors in students and to have them act intentionally. In addition to this objective, education also seeks to determine deficiencies in students' learning, to correct any mistakes in their knowledge, and to provide effective feedback to students. Accordingly, students have at their disposal practical information to support their learning like what goals and proficiencies they have completed or acquired, and what skills they still need to attain. It is important that students' goals and learning outcomes are met in order to carry out the education process in a qualified way. Undoubtedly, one of the important components in this process is evaluation. The quality of the education system, its functioning and whether the process is proceeding effectively can only be determined by evaluation. Although evaluation is carried out at the end of the process and is seen as independent of teaching, it provides information on many variables such as how students prepare for lessons, what they experience with the learning process and how they prepare for the exam. Evaluations are used by educational systems not only to determine students' vocational or creative inclinations but also to increase teaching quality (Turgut & Baykul, 2010). Evaluations are needed for a variety of purposes in education systems, including diagnostic, formative and summative types.

Diagnostic evaluations are conducted to determine whether potential students have the necessary behaviors and/or cognitive skills required to attend a program, course, or activity (Tekin, 2005). This type of evaluation seeks to measure students' pedagogical characteristics as opposed to psychological ones (Özcelik, 2010). Formative-type evaluations are conducted after every unit or topic to determine students' deficiencies and to provide feedback related to these deficiencies. Consequently, formative-type evaluations are considered a part of the greater instruction process. Instead of gauging students' academic performance, this type of evaluation is given during the course of the instruction process to determine what aspects of a unit or topic students could improve upon and what their gaps in knowledge are. Summative evaluations are yet another type of evaluation that encompasses more than one proficiency and are given either during or after the instruction process. This type of evaluation seeks to assign a grade to students, to determine whether they have a sufficient foundation to continue to the subsequent step in their course or program, and serve as a predictor of future performance (Tekin, 2005). Accordingly, the degree to which students have attained the goals and

proficiencies expected of them and how prepared they are to proceed to the subsequent stage in their education may be measured.

Evaluations should not be considered independent from or as a tool administered solely at the end of instruction. Evaluations present information on different variables like how prepared students are for tests, what difficulties they encountered during instruction, and how well they have studied their lessons (Birenbaum, 1997; Struyven, Dochy & Janssens, 2005). According to Biggs (2003), evaluations make significant contributions to students' learning progress. For this reason, a quality evaluation method should not only constitute an integral part of students' learning processes but also guide students on how to make the most of their learning during this process (Gulbahar & Buyukozturk, 2008).

Evaluation preferences are defined as scoring procedures used to measure students' academic performance (Birenbaum, 1994). Both traditional and supplementary evaluations are frequently used while performing in-class evaluations to measure students' academic performance. Traditional evaluation approaches generally measure academic performance outside the normal flow of instruction, are based on output, and frequently use such items as multiple-choice, short-answer, true-false, matching, and fill-in-the-blank type items. Seeking primarily to measure the level of students' cognitive abilities to recall and conceptualize, this type of evaluation is preferred by teachers who adopt a constructive approach. In fact, most teachers consider themselves proficient in this type of evaluation approach (Gelbal & Kelecioglu, 2007). Supplementary evaluation approaches like performance-based tasks, portfolios, and rubrics seek to measure students' higher-level cognitive abilities (e.g., critical thinking, researching, creativity) are used to assess educational goals (Kutlu, Dogan & Karakaya, 2008). One of the most important differences between supplementary and traditional evaluation approaches is that the former allows one to assess educational goals in addition to measuring individual differences and multiple intelligence areas.

The evaluation method that students prefer to during instruction may influence their academic performance (Biggs, 2003). The literature contains several studies (Biggs, 2003; Birenbaum, 2003; Ekinci, 2009; Struyven et al., 2005) discussing how students' preferences regarding instructional and evaluation methods impact their academic performance, their perceptions related to learning, and how they participate in learning processes and then, how all of these work together to impact students' evaluation preferences. Previous studies have found that there are significant differences in students' academic performance and preferred type of evaluations based on their learning approaches (Mayya, Rao & Ramnarayan, 2004; McManus, Richards & Winder, 1999; Sambell et al., 1997; Scouller, 2000). Just as students have different intelligence areas, they also have different learning approaches. While some students are familiar with practice-based learning methods, others may prefer teaching methods based entirely on direct instruction. Students' learning habits may cause them to adopt specific test preparation habits, which, in turn, may cause them to prefer specific evaluation types over others. A review of the literature, however, reveals that the majority of studies focus more on statistical significance and

relationship with regard to students' evaluation preferences and that there are only a few studies (Altun & Gelbal, 2014; Birgin & Gürbüz, 2008; Gelbal & Kelecioglu, 2007; Gijbels & Dochy, 2006; Isnac, 2018; Struyven, Dochy & Janssens, 2005; Struyven, Dochy & Janssens, 2005; Sahin, Ozturk Boztunc & Teker Tasdelen, 2015) seeking to identify what type of evaluations students like to take or which types they prefer over others. Of these studies, only in Isnac (2018) and Sahin, Ozturk Boztunc and Teker Tasdelen (2015) are scaling methods used to assess students' evaluation preferences. In their study, Sahin, Ozturk Boztunc and Teker Tasdelen (2015) found that pre-service teachers preferred true-false tests the most and performance-based tasks the least. Isnac (2018) conducted a study with middle school students in which she scaled students' evaluation preferences according to their learning approaches. Since in scaling approaches, psychological and emotional characteristics can be scaled and a shared point regarding individuals' preferences can be reached (Anil & Guler, 2006; Kan, 2008; Ozkan Ozer & Guvendir Acar, 2011), one method to determine which type of evaluation students prefer is to scale their preferences after having asked them directly. An examination of the literature reveals that only a single study exists investigating university students' evaluation preferences using rank-order judgment scaling. Accordingly, the current study sought to identify which types of evaluations university students preferred, to scale their preferences according to rank-order judgments, and to determine scale values for their preferences. To accomplish this objective, responses to the following questions were solicited:

1. What evaluation methods do university students prefer to measure their academic performance?

2. Using rank-order scaling, what are the item scale values for university students' evaluation preferences?

Method

Research Design

The survey model used in this study aims to describe a situation that has existed or still exists (Cresswell, 2003; Karasar, 2014). Accordingly, the current study follows a survey model since university students' evaluation preferences were sought.

Research Sample

The study group was described in this research. The study was conducted with a total of 376 students enrolled in the faculty of education in two different universities in Turkey (i.e., Canakkale Onsekiz Mart University and Trakya University). Approximately 31% (n=116) of the students participating in the study were male and 69% (n=260) were female. Table 1 presents student distributions by university and department.

Table 1

Student Distribution by Department

Department	f	%
German Language Education	11	2.9
Computer and Instructional Technologies	38	10.1
English Language Education	64	17.0
Japanese Language Education	24	6.4
Psychological Counseling and Guidance	136	36.2
Art Education	16	4.3
Social Sciences Education	35	9.3
Turkish Language Education	52	13.8
Total	376	100.0

Of the participating students, 2.9% were enrolled in the German Language Education Department, 10.1% in the Computer and Instructional Technologies Department, 17% in the English Language Education Department, 36.2% in the Psychological Counseling and Guidance Department, 4.3% in the Art Education Department, 9.3% in the Social Sciences Education Department, and 13.8% in the Turkish Language Education Department. Table 2 presents student distributions by grade level.

Table 2

Student Distribution by Grade

Grade level	f	%
1.00	56	14.9
2.00	109	29.0
3.00	203	54.0
4.00	8	2.1
Total	376	100.0

As seen in Table 2, 14.9% of the participating students were in their first year, 29% in their second year, 54% in their third year and 2.1% in their fourth year of undergraduate education.

Research Instruments and Procedures

The data collection tool was developed by the researcher for this study. During the development of the data collection tool, the researcher performed a review of the literature and determined potential tools that may be used to assess students' academic performance. A semi-structured interview form was also used to solicit

responses regarding students' evaluation preferences. A total of 80 students enrolled in different grade levels within the faculty of education completed this interview form in which they were asked to rank their evaluation preferences in writing. After examining students' responses to the interview form, similar responses were aggregated to form a total of 16 items. The items were subsequently presented to three measurement and evaluation experts whose opinions were used to construct a 13-item tool measuring students' evaluation preferences. The 13 items included in the measurement tool were: (i) preparing individual presentations, (ii) taking frequent quizzes, (iii) doing a project as homework instead of a test, (iv) taking an open-book and notes test (no time limit), (v) answering open-ended questions (closed book and notes, with time limit), (vi) doing group homework instead of a test, (vii) taking an oral exam, (viii) doing individual homework, (ix) answering short-answer and open-ended questions, (x) taking a multiple-choice test, (xi) taking a test including different types of questions simultaneously (e.g., multiple-choice, short-answer, true-false, matching), (xii) doing individual homework, and (xiii) taking an open-book and notes test (with time limit). The researcher implemented the measurement tool with the students, which took approximately 5 minutes. The participants were provided with all necessary explanations regarding the measurement tool and how to complete it prior to its administration.

Data Analysis

Rank-order judgments based on Thurstone's (1927) law of comparative judgment (Turgut & Baykul, 1992) were used to analyze the data collected for the study. The rank-order judgment scaling method is itself based on one of two fundamental scaling approaches (i.e., the judgment and reaction). In the judgment approach, stimuli are scaled according to observers' judgments in a predetermined dimension. This approach seeks to define the degree of stimulation of K number of stimuli for each of the N number of observers. In this approach, the observer's duty is to rank each stimulus in the scaling dimension in reference to the other stimuli. Accordingly, the average of observer judgments for any stimulus constitutes its scale value. The reaction approach, however, determines individuals' reactions by executing K number of stimuli to a group composed of N number of individuals. In this approach, the individuals reacting are not objective experts. They only determine the position of the stimulus examined in reference to the other stimuli on the same scale (Anil & Guler, 2006; Guilford, 1954; Tezbasaran, 2004; Torgerson, 1958). In the this study, rank-order judgment scaling based on the judgmental decisions approach was used.

Since ranking is based on the size difference between stimuli, it resembles a pair-wise comparative approach. Moreover, since participants experience fewer contradictions in rank-order judgment scaling, the results are potentially more consistent (Guilford, 1954; Turgut & Baykul, 1992). All stimuli are given to observers in rank-order judgment scaling. Observers rank stimuli by assigning each stimulus a rank number. The basic premise in this scaling approach is that each stimulus is given a rank number. This way, the entire stimuli group is transformed into a uniform standard to which each individual stimulus is compared. Scale values are attained by comparing the ratio attained from the rank-order judgments assigned to stimuli with the uniform standard. Subsequently, the pair-wise comparison is analyzed as in the fifth equation of Thurstone's (1927) law of comparative judgment theory (Anil & Inal,

2018; Guilford, 1954). In this study, a rank-order frequencies matrix was constructed showing how many times and in which order students placed each of their preferred type of evaluation. A ratios matrix was constructed using the rank-order judgments that students assigned to their evaluation preferences. The unit normal deviation matrix was constructed by calculating the z-values corresponding to the ratio's matrix elements. The total of each column's values was written at the bottom row of the unit normal deviation matrix. Scale values were then attained by calculating the average of each z-value appearing in this row for each separate column.

Results

This section presents students' rankings for evaluation preferences in regard to scaling procedure steps and explanations. In order to answer the first secondary objective of this study, a rank-order frequencies matrix was obtained using the rankings of 376 students' preferences regarding 13 stimuli.

Table 3
Frequencies Matrix

Rank-Order Frequencies Matrix for Stimuli														
R	**r**	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**Total**
1	13	17	55	9	55	28	28	82	3	5	5	56	7	26	**376**
2	12	16	58	14	30	41	16	62	11	20	11	32	19	46	**376**
3	11	19	35	15	29	51	25	47	14	30	19	41	15	36	**376**
4	10	22	41	30	19	51	26	27	24	28	16	33	26	33	**376**
5	9	22	30	15	27	39	29	42	28	34	15	32	27	36	**376**
6	8	30	27	27	26	34	24	32	31	35	23	27	33	27	**376**
7	7	22	17	35	19	26	28	25	34	43	37	33	28	29	**376**
8	6	38	23	36	26	27	38	10	30	36	28	24	35	25	**376**
9	5	45	20	41	20	19	30	18	42	24	41	24	23	29	**376**
10	4	29	23	39	24	19	31	14	49	37	28	33	40	10	**376**
11	3	32	18	43	19	20	39	11	38	38	50	22	23	23	**376**
12	2	38	14	42	22	16	30	4	42	31	52	10	32	43	**376**
13	1	46	15	30	60	5	32	2	30	15	51	9	68	13	**376**
	Total	376	376	376	376	376	376	376	376	376	376	376	376	376	

Table 3 consists of a frequencies table depicting the rank numbers assigned by 376 judgments for each of the 13 total stimuli. The total number of observers for each row and column is 376. Following this procedure, n(sjk>ski) tables for all stimuli were prepared and a pair-wise frequencies matrix for all stimuli was constructed.

The ratios matrix in Table 4 was constructed by dividing the column totals in the frequency matrix by the square number of total participants ($n^2=376^2$).

Table 4

Ratios Matrix for Evaluation Preferences

	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
A		0.30	0.51	0.42	0.30	0.45	0.21	0.52	0.44	0.56	0.32	0.52	0.37
B	0.70		0.66	0.55	0.47	0.60	0.37	0.67	0.61	0.70	0.48	0.66	0.53
C	0.49	0.34		0.41	0.28	0.44	0.19	0.51	0.42	0.56	0.30	0.51	0.36
D	0.58	0.45	0.59		0.42	0.54	0.33	0.60	0.54	0.63	0.42	0.60	0.47
E	0.70	0.53	0.72	0.58		0.65	0.38	0.73	0.66	0.76	0.50	0.71	0.56
F	0.55	0.40	0.56	0.46	0.35		0.26	0.57	0.49	0.61	0.36	0.57	0.42
G	0.79	0.63	0.81	0.67	0.62	0.26		0.82	0.76	0.84	0.62	0.80	0.67
H	0.48	0.33	0.49	0.40	0.27	0.43	0.18		0.42	0.55	0.29	0.50	0.34
I	0.56	0.39	0.58	0.46	0.34	0.51	0.24	0.58		0.63	0.36	0.58	0.41
J	0.44	0.30	0.44	0.37	0.24	0.39	0.16	0.45	0.37		0.26	0.46	0.31
K	0.68	0.52	0.70	0.58	0.50	0.64	0.38	0.71	0.64	0.74		0.70	0.55
L	0.48	0.34	0.49	0.40	0.29	0.43	0.20	0.50	0.42	0.54	0.30		0.35
M	0.63	0.47	0.64	0.53	0.44	0.58	0.33	0.66	0.59	0.69	0.45	0.65	

In addition to the ratios matrix presented in Table 4, the analysis continued using scaling methods based on a pair-wise comparative approach. In ratios matrices, the most significant point diagonals 1 was given to the most important point in the ratios matrix. Upon examination of Table 4, it is observed that the sum of diagonals for the ratios matrix was equal to 1. After confirming this, the analysis continued using the fifth equation like in the pair-wise comparative scaling method. As a result, a unit normal deviation matrix (z) was constructed and presented in Table 5.

Table 5
Unit Normal Deviation Matrix for Evaluation Preferences

	z Matrix (unit normal deviation matrix)												
	A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**
A		-0.53	0.02	-0.21	-0.52	-0.12	-0.79	0.04	-0.16	0.15	-0.48	0.05	-0.34
B	0.53		0.42	0.12	-0.07	0.26	-0.33	0.45	0.28	0.53	-0.06	0.42	0.07
C	-0.02	-0.42		-0.23	-0.57	-0.14	-0.86	0.02	-0.19	0.14	-0.52	0.03	-0.37
D	0.21	-0.12	0.23		-0.20	0.10	-0.44	0.25	0.10	0.34	-0.19	0.25	-0.08
E	0.52	0.07	0.57	0.20		0.38	-0.32	0.62	0.41	0.70	0.01	0.57	0.15
F	0.12	-0.26	0.14	-0.10	-0.38		-0.65	0.17	-0.03	0.27	-0.35	0.17	-0.21
G	0.79	0.33	0.86	0.44	0.32	-0.65		0.92	0.71	0.99	0.30	0.84	0.44
H	-0.04	-0.45	-0.02	-0.25	-0.62	-0.17	-0.92		-0.21	0.13	-0.56	0.01	-0.40
I	0.16	-0.28	0.19	-0.10	-0.41	0.03	-0.71	0.21		0.33	-0.36	0.20	-0.22
J	-0.15	-0.53	-0.14	-0.34	-0.70	-0.27	-0.99	-0.13	-0.33		-0.65	-0.10	-0.49
K	0.48	0.06	0.52	0.19	-0.01	0.35	-0.30	0.56	0.36	0.65		0.52	0.13
L	-0.05	-0.42	-0.03	-0.25	-0.57	-0.17	-0.84	-0.01	-0.20	0.10	-0.52		-0.39
M	0.34	-0.07	0.37	0.08	-0.15	0.21	-0.44	0.40	0.22	0.49	-0.13	0.39	
Total	2.89	-2.63	3.14	-0.44	-3.87	-0.18	-7.59	3.50	0.95	4.81	-3.53	3.36	-1.71
Mean	0.24	-0.22	0.26	-0.04	-0.32	-0.02	-0.63	0.29	0.08	0.40	-0.29	0.28	-0.14
SJ	0.87	0.41	0.89	0.60	0.31	0.62	0.00	0.92	0.71	1.03	0.34	0.91	0.49

In order to determine university students' evaluation preferences, the data collected by the measurement tools were scaled based on rank-order judgments. The smallest value on the unit normal distribution matrix was -7.59 for evaluation preference G. By taking this value as the ranking criteria, an absolute value of 7.59 was attained and is shown on a number line in Figure 1.

| 0,87; A | 0,41; B | 0,89; C | 0,60; D | 0,31; E | 0,62; F | 0,00; G | 0,92; H | 0,71; İ | 1,03; J | 0,34; K | 0,91; L | 0,49; M |

Figure 1. Scale values placed on a number line

Table 6

Evaluation Preferences and Scale Values

Stimulus Rank Number	Scale Values	Evaluation Preference
9	0.87	Preparing Individual presentations.
4	0.41	Taking frequent quizzes.
10	0.89	Doing a project as homework instead of a test.
6	0.60	Taking an open-book and notes test (no time limit).
2	0.31	Answering open-ended questions (closed book and notes, with a time limit).
7	0.62	Doing group homework instead of a test.
1	0.00	Taking an oral exam.
12	0.92	Doing individual homework.
8	0.71	Answering short-answer and open-ended questions.
13	1.03	Taking a multiple-choice test.
3	0.34	Using a given individual article to write as homework.
11	0.91	Taking a test, including different types of questions simultaneously (e.g., multiple-choice, short-answer, true-false, matching).
5	0.49	Taking an open-book and notes test (with a time limit).

As seen in Table 6, the most preferred evaluation type by students was taking an oral exam whereas the least preferred was taking a multiple-choice test. Students' evaluation preferences from most to least preferred were (1) taking an oral exam, (2) answering open-ended questions (closed book and notes, with time limit), (3) using a given individual article to write as homework, (4) taking frequent quizzes, (5) taking an open-book and notes test (with time limit), (6) taking an open-book and notes test (no time limit), (7) doing group homework instead of a test, (8) answering short-answer and open-ended questions, (9) preparing Individual presentations, (10) Doing a project as homework instead of a test, (11) taking a test including different types of

questions simultaneously (e.g., multiple-choice, short-answer, true-false, matching), (12) doing individual homework, and (13) taking a multiple-choice test.

Discussion, Conclusion, and Recommendations

Using rank-order judgments-based scaling (one of the scaling methods based on judgment decisions), this study has sought to determine what evaluation types that university students enrolled in different departments preferred to be used to measure their academic performance.

This study found that students most preferred oral exams to measure their academic performance. Oral exams are defined as a type of evaluation in which questions are generally asked and responses are given orally (Tekin, 2005). Oral exams have historically been frequently used in educational environments. Although students are less likely to suffer from nervousness and unnecessary fear while taking oral exams, the existence of subjectivity in the scoring system, the interaction between the examiner and examinee, the preparation of individual questions for each examinee, and the fact that they need to be conducted individually render them considerably difficult to administer (Turgut & Baykul, 2010). Other hurdles that render the administration of oral exams even more difficult include the need to record answers given during the test and the need to ensure proper oversight of the exam in order to ensure impartiality and to avert any potential legally questionable situations (Sezer & Bilgin, 2009).

Students' second most preferred evaluation type was answering open-ended questions which they could use their book and notes to respond and for which a time limit was set. Although open-ended questions are appropriate to measure high-level cognitive skills, there are some difficulties in scoring them (e.g., scoring not being objective, low content validity) (Atilgan, Kan & Dogan, 2009). In their study examining the appropriateness of open-ended questions used in tests prepared by middle school teachers, Incecam, Demir and Demir (2018) found that teachers made extensive use of open-ended questions. The results of a study by Sahin, Ozturk Boztunc and Teker Tasdelen (2015) reveal that open-ended questions were the fourth most preferred method of evaluation by pre-service teachers. Scouller (1998) found that the majority of students who adopt a deep learning approach performed better on exams that included written open-ended item format.

The findings of the current study revealed that the third most preferred evaluation type by university students was using a given individual article to write as homework. Individual learning and wanting to complete the homework given to oneself are considered general characteristics of students espousing a deep learning approach (Minbashian, Huon & Bird, 2004; Prosser & Trigwell, 1999; Ramsden, 1991). In his study conducted with university students enrolled in a faculty of education, Scouller (1998) found that while those students embracing a deep learning approach performed poorly on multiple-choice item format, they received higher scores on written tests. Accordingly, students' preferences in being given an article to write as homework may be considered a sign that they have adopted a deep learning approach. As such, the findings of the current study are consistent with those of Scouller's (1998).

Examination of students' three least preferred evaluation types revealed that they preferred stimulus to be assessed by multiple-choice type questions. Despite nearly all of the large-scale tests students have taken from elementary school to university are multiple-choice tests, the fact that this type of test was the least preferred by students was an important finding of this study. This specific finding contradicts those obtained by Struyven, Dochy, and Janssens (2005) study, in which students stated that they experienced less anxiety and had higher expectations for success when responding to multiple-choice questions that did not require them to construct their own answers.

Students' second least preferred type of evaluation was individual homework. Students stated either that they did not like being assessed by individual homework or that they made efforts to avoid it altogether. Similarly, Ozer-Ozkan and Acar-Guvendir (2013) conducted a scaling study with students enrolled in a measurement and evaluation class in which they attempted to identify students' preferred teaching styles. In their study, they found that students' least preferred evaluation types were narratives and individual work. In a similar vein, Cross (1981) asserted that individual work could cause deficiencies in communication between student and instructor. Furthermore, students' fears of being subjected to a subjective grading scale may cause them to dislike being assessed by individual homework.

Students' third least preferred type of evaluation were those composed of different types of questions (e.g., multiple-choice, short-answer, true-false, matching). The nation-wide tests that students take (e.g., Council of Higher Education Exam [YKS], High School Entrance Exam [LYS]) are composed strictly of multiple-choice questions and include no other types of questions. The fact that students are most familiar with tests composed of a single type of question may explain their dislike of being assessed by tests containing a multitude of question types.

Several recommendations may be made based on the findings of this study. The first recommendation is that student anxiety may be slightly reduced by explaining them how the evaluation will be structured and by providing them a grading rubric when they are to be given open-ended questions or homework. As stated by the majority of students, a single type of question may be used on tests or students can be briefed on other question types prior to being given a test composed of more than one question type in order to reduce their prejudices toward those with which they are less familiar. Recommendations for subsequent studies include asking students to provide qualitative explanations as to why they ranked their preferred question types in evaluations in the order they did, as doing so will shed light on the rationale underlying their choices. Since this study was restricted to university students enrolled in the faculty of education of two universities, other researchers may conduct similar studies with students enrolled in other faculties in the same or different higher education institutions, and compare the results between them.

References

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*, 5(1), 1-11.

Anil, D. & Guler, 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.

Anil, D. & Inal, H. (2017). *Psikofizikte ölçekleme uygulamaları*. Ankara: PegemAkademi.

Atilgan, H., Kan, A. & Dogan, N. (2007). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı Yayıncılık.

Biggs, J. (2003). *Teaching for quality learning at university* (2nd edition). United Kingdom: The Society for Research into Higher Education and Open University Pres.

Birenbaum, M. (1994). Toward adaptive evaluation - the student's angle. *Studies in Educational Evaluation*, 20, 239-255.

Birenbaum, M. (1997). Evaluation preferences and their relationship to learning strategies and orientations. *Higher Education*, 33, 71-84.

Birenbaum, M. (2003). New insights into learning and teaching and their implications for evaluation. Segers, M., Dochy, F. ve Cascallar, E. (eds.). *Optimizing New Methods of Evaluation: In Search of Qualities and Standards* (13-36). Dordrecht, The Netherlands: Kluwer.

Birgin, O., & Gürbüz, R. (2008). Sınıf öğretmeni adaylarının ölçme ve değerlendirme konusundaki bilgi düzeylerinin incelenmesi [Investigation of pre-service primary teachers' knowledge level about measurement and evaluation]. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 20, 163-179.

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory.* Harcourt Brace Jovanovich College Publishers: Philadelphia.

Cross, K. P. (1981). *Adults as learners.* San Francisco: Jossey-Bass.

Ekinci, N. (2008). *Üniversite öğrencilerinin öğrenme yaklaşımlarının belirlenmesi ve öğretme öğrenme süreci değişkenleri ile ilişkisi*. (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi, Ankara.

Gelbal, S. & Kellecioglu, H. (2007). Öğretmenlerin ölçme ve değerlendirme yöntemleri hakkındaki yeterlik algıları ve karşılaştıkları sorunlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 33, 135-145* .

Gijbels, D. & Dochy, F. (2006). Students' evaluation preferences and approaches to learning: can formative evaluation make a difference?. *Educational Studies*, 32(4), 399-409.

Guilford, J. P. (1954). *Psychometrics methods*. New York: Mc Graw-Hill Book Co.

Gulbahar, Y. & Buyukozturk, Ş. (2008). Değerlendirme tercihleri ölçeğinin Türkçeye uyarlaması [Adaptation of evaluation preferences inventory to Turkish]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 35, 148-161.

Incecam, B., Demir, E. & Demir, E. (2018). Ortaokul öğretmenlerinin sınıf içi ölçme ve değerlendirmelerde yazılı yoklamalarda kullandıkları açık uçlu maddeleri hazırlama yeterlikleri [Competencies of middle school teachers to prepare open-ended items used in open-ended test for in-classroom evaluation]. *Elementary Education Online*, 17(4): 1912-1927

Isnac, F. (2018). *Ortaöğretim öğrencilerinin öğrenme yaklaşımlarına göre değerlendirme tercihlerinin ölçeklenmesi* [Scaling evaluation preferences of secondary school students according to learning approaches]. Master Thesis, Ankara University, Educational Sciences Institute, Ankara.

Kan, A. (2008). Psikolojik değişkenleri ölçmek için kullanılan ölçekleme yaklaşımları üzerine bir karşılaştırma [A comparison between scaling procedures for measuring psychological variables]. *Eğitimde Kuram ve Uygulama*, 4(1), 2-18.

Karasar, N. (2014). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayın Dağıtım.

Kutlu, O., Dogan, C. D. & Karakaya, I. (2008*). Öğrenci başarısının belirlenmesi performansa ve portfolyoya dayalı durum belirleme*. Ankara: Pegem Akademi yayıncılık.

Mayya, S., Rao, A.K. & Ramnarayan, K. (2004). Learning approaches, learning difficulties and academic performance of undergraduate students ofphysiotherapy. *The Internet Journal of Allied Health Sciences and Practice*, 2(4), 1-6.

McManus, I.C., Richards, P. & Winder, B.C. (1999). Intercalated degrees, learning styles, and career preferences: Prospective longitudinal study of UK medical students. *British Medical Journal International edition*, 319, 542-546.

Minbashian, A., Huon, G. F. & Bird, K. D. (2004). Approaches to studying and academic performance in short-essay exams. *Higher Education*, 47,161-176.

Ozcelik, D. A. (2011). *Ölçme ve değerlendirme*. Ankara: Pegem Akademi

Ozkan Ozer, Y. & Guvendir Acar, M. (2013). Öğrencilerin ölçme ve değerlendirme dersinin sunulmasında tercih ettikleri öğretim yöntemleri [Students' teaching methods preferences for measurement and evaluation course]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi,* 4(1), 1-14.

Prosser, M. & Trigwell, K. (1999). Relational perspectives on higher education teaching and learning in the science. *Studies in Secience Education*, 33(1), 31-60.

Ramsden, P. (1991). A performance indicator of teaching quality in higher education: The Course Experience Questionnaire, *Studies in Higher Education*, 16:2, 129-150.

Sambell, K., McDowell, L. & Brown, S. (1997). 'But is it fair?': An exploratory study of student perceptions of the consequential validity of evaluation. *Studies in Educational Evaluation,* 23(4), 349-371.

Scouller, K. (2000). The influence of evaluation method on student learning. *Australian Association for Research in Education*, 35(4), 453-472.

Scouller, K. (1998). The influence of evaluation method on students' learningapproaches: Multiple choice question examination versus assignment essay. *Higher Education*, 35,   453–472.

Sezer, Y. & Bilgin, H. (2009). Sözlü sınavların yargısal denetimi [Judicial review of oral exams]. *TBB Dergisi*, 86, 168-187.

Struyven, K., Dochy, F. & Janssens, S. (2005). Students' perceptions about evaluation and evaluation in higher   education: A review. *Evaluation & Evaluation in Higher Education*, 30(4), 325-341.

Sahin, M. G., Ozturk Boztunc N. & Teker Tasdelen, G.(2015). Öğretmen adaylarının başarılarının değerlendirilmesinde tercih ettikleri ölçme araçlarının belirlenmesi [Determining the pre-service teachers' measurement tool preferences for   evaluation of   their achievement]. *Eğitimde ve Psikolojide Ölçme   ve Değerlendirme Dergisi*, 6 (1), 95-106.

Tekin, H. (2004). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı yayın evi.

Tezbasaran A. (2004). Likert tipi ölçeklere madde seçmede geleneksel madde analizi tekniklerinin karşılaştırılması. *Türk Psikoloji Dergisi*, 19 (54), 77-90.

Thurstone, L. L. (1927). A law of comparative judgement. *Psychological Rewiev*, 34, 273-286.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: John Wiley & Son.

Turgut, M. F. & Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi

Turgut, F., M., & Baykul, Y. (1992). *Ölçekleme teknikleri.* Ankara: ÖSYM Yayınları

Öğrencilerin Değerlendirme Tercihlerinin Sıralama Yargılarına Dayalı Ölçeklenmesi

Özet

Problem Durumu: Eğitim sürecinin önemli ögelerinden biri değerlendirmedir. Eğitim sisteminin, işleyişi ve sürecin etkili bir şekilde ilerleyip ilerlemediğinin kontrolü ancak

değerlendirme ile belirlenebilmektedir. Ölçme sonuçlarının belirlenen ölçüt veya ölçütler takımıyla karşılaştırılarak bir karara varma süreci olan değerlendirme, öğrencilerin mesleki veya beceri açıdan eğilimlerinin belirlenmesine yardımcı olmasının yanısıra öğretimin niteliği hakkında ilgili paydaşlara önemli bilgiler sağlar (Turgut & Baykul, 2010). Değerlendirme, sadece öğrenme sürecinin sonunda gerçekleştirilen ve öğretimden bağımsız bir durum olarak ele alınamaz. Aynı zamanda öğrenci başarısı, hedef ve kazanımlara ulaşma düzeyleri hakkında da bilgi sağlar. Ayrıca, öğrencilerin sınavlara nasıl hazırlandıklarını, öğrenme sürecinde karşılaştıkları zorlukları, derslerine nasıl çalıştıkları gibi farklı değişkenler hakkında da bilgi elde edilmesine yardımcı olur (Birenbaum, 1997; Struyven, Dochy & Janssens, 2005). Biggs (2003)'e göre, değerlendirme öğrenmenin gelişmesine önemli katkı sunmaktadır. Bu nedenle de, nitelikli bir değerlendirme yöntemi, öğrencilerin öğrenme süreçlerinin bir parçası olmalı ve öğrencilere öğrenme sürecinde nasıl öğrenmeleri gerektiği konusunda da rehberlik etmelidir (Gulbahar & Büyüköztürk, 2008).

Öğrencilerin öğretimsel tercihleri ve değerlendirme yöntemlerine ilişkin tercihlerinin, akademik başarı, öğrenmeye ilişkin algıları ve öğrencilerin öğrenme sürecini nasıl gerçekleştirdiklerini ve de tüm bunların değerlendirme tercihine nasıl bir etkiye sahip olduğuna ilişkin alan yazında (Biggs, 2003; Brenbaum, 2003; Struyven & Digerleri, 2005; Ekinci, 2009) yapılmış araştırmalar bulunmaktadır. Yapılan araştırmalarda (Mayya, Rao ve Ramnarayan, 2004; McManus, Richards & Winder 1999; Sambell, McDowell & Brown, 1997; Scouller, 2000) öğrenme yaklaşımlarının öğrencilerin akademik başarıları ile değerlendirme tercihlerine anlamlı düzeyde farklılık oluşturduğu tespit edilmiştir. Ancak alan yazın incelendiğinde, yapılan araştırmaların çoğunluğunun, öğrencilerin değerlendirme tercihlerine ilişkin daha çok istatistiksel anlamlılık ve ilişki üzerine olduğu tespit edilmiş (Gelbal & Kelecioglu, 2007; Birgin & Gurbuz, 2008; Struyven, Dochy & Janssens, 2005; Gijbels & Dochy, 2006); öğrencilerin ne tür değerlendirme türlerini tercih ettiklerini ortaya çıkaran az sayıda (Altun ve Gelbal, 2014; Isnac, 2018; Sahin, Ozturk Boztunc & Teker Tasdelen, 2015) araştırma bulunmaktadır. Bu araştırmalardan sadece Isnac (2018) ve Sahin, Ozturk Boztunc & Teker Tasdelen (2015) tarafından yapılan araştırmalarda, öğrencilerin değerlendirme tercihleri ölçekleme yöntemleriyle belirlenmiştir. Alan yazın incelendiğinde, üniversite öğrencilerinin değerlendirme tercihlerinin sıralama yargılarına dayalı ölçekleme yoluyla incelendiği sadece bir çalışmaya (Sahin, Ozturk Boztunc & Teker Tasdelen, 2015) rastlanmıştır. Öğrencilerin başarılarının değerlendirilmesinde tercih ettikleri değerlendirme türlerinin belirlenmesi, öğrencilerin çalışma alışkanlıkları ve sınavlara nasıl hazırlandıkları hakkında önemli bilgiler ortaya koyacağı düşünülmektedir. Bu bağlamda araştırmanın problemi, üniversite öğrencilerinin değerlendirme tercihlerine ilişkin ölçek değerlerinin belirlenmesidir.

Araştırmanın Amacı: Araştırmanın amacı, üniversite öğrencilerinin değerlendirme tercihlerinin belirlenmesi ve değerlendirme tercihlerinin sıralama yargılarına göre ölçeklenmesidir.

Araştırmanın Yöntemi: Araştırma tarama modeli ile tasarlanmıştır. Araştırma Canakkale Onsekiz Mart ve Trakya Üniversitesi eğitim fakültelerinde öğrenim görmekte olan toplam 376 üniversite öğrencisi ile gerçekleştirilmiştir. Araştırmada veri toplama aracı araştırmacı tarafından geliştirilmiştir. Veri toplama aracının

geliştirme aşamasında, alan yazın taranmış ve öğrencilerin akademik başarılarını değerlendirmede kullanılabilecek ölçme araçları belirlenmiştir. Ayrıca araştırma kapsamında öğrencilere değerlendirme tercihlerinin sorulduğu yarı yapılandırılmış bir görüşme formu kullanılmıştır. Bu görüşme formu, eğitim fakültesinde çeşitli sınıflarda öğrenim görmekte olan 80 öğrenciye uygulanmış ve değerlendirme tercihlerini sırasız olarak yazmaları istenmiştir. Öğrencilerin vermiş oldukları yanıtlar incelenerek ortak olanlar (toplam 16 madde) bir araya getirilmiştir. Daha sonra bu maddeler uzmanların (üç ölçme ve değerlendirme uzmanı) görüşüne sunulmuş ve uzman görüşleri doğrultusunda 13 madde ölçme aracının maddelerini oluşturmuştur.

Araştırmanın Bulguları: Bulgularda öncelikle toplam 13 uyarıcı için 376 yargıcının her bir uyarıcı için vermiş oldukları sıra numaralarına ilişkin frekans tablosu oluşturulmuştur. Satır ve sütunların toplam gözlem sayısı 376'dır. Bu işlemden sonra ikili olarak tüm uyarıcılar arasındaki $n(s_{jk}>s_{ki})$ tabloları hazırlanmış ve tüm uyarıcılar arasında ikili sıra frekanslar matrisi elde edilmiştir. Daha sonra oranlar matrisi hesaplanmıştır. Oranlar matrisinde en önemli nokta köşegenleri toplamının 1'i vermesidir. Buna göre, oranlar matrisindeki köşegenler toplamının 1'e eşit olup olmadığı incelenmiş ve köşegenler toplamının 1'e eşit olduğu tespit edilmiştir. Bu kontrolün ardından, V. Hal denklemi kullanılarak, ikili karşılaştırma yöntemi ile analize devam edilmiştir. Daha sonra birim normal sapmalar matris değerleri hesaplanmıştır.

Sonuç ve Öneriler: Araştırmadan elde edilen bulgulara göre, öğrenciler en çok sözlü sınavı tercih etmişlerdir. Sözlü sınav her ne kadar, öğrenciyi heyecandan ve sınav ile ilgili gereksiz korkulardan kurtarsa da, puanlanmasındaki öznellik, sınav yapan ile yapılan arasındaki etkileşim ve sınav sorularının her birey için önceden hazırlama gerekliliği gibi sebeplerden dolayı kalabalık gruplarda gerçekleştirilmesi oldukça zor olabilmektedir. (Turgut & Baykul, 2010).

Öğrencilerin değerlendirme türü olarak en çok tercih ettiği ikinci uyarıcı ise "sınırlı süre içerisinde, kitap ve defter açık bir şekilde açık uçlu sorularla değerlendirme" yapılmasıdır. Açık uçlu sorular ölçtüğü özellik açısından üst düzey bilişsel beceriler ölçmek için uygun olsa da, soruların değerlendirilmesi ve puanlanmasındaki bazı sorunlar (puanlamanın objektif olmaması, kapsam geçerliğinin düşük olması gibi) taşımaktadır (Atılgan, Kan & Dogan, 2009). Scouller (1998) tarafından yapılan araştırma sonuçlarına göre, derin öğrenme yaklaşımına sahip öğrencilerin çoğunlukla yazılı (açık uçlu) yoklama soru türündeki sınavlarda daha başarılı oldukları belirtilmiştir.

Araştırma sonuçlarına göre, öğrencilerinin tercih ettiği üçüncü değerlendirme türü ise "Bireysel makale ödevi verilerek" yapılan değerlendirmedir. Bireysel öğrenme ve verilen ödevi istekli olarak yerine getirme davranışı, çoğunlukla derin öğrenme yaklaşımını benimseyen öğrencilerin genel özelliklerinden biridir (Minbashian, Huon & Bird, 2004; Prosser & Trigwell, 1999; Ramsden, 1991). Bu bağlamda, araştırma grubunda yer alan ve derin öğrenme özelliğine sahip öğrenciler, bireysel makale ödevini değerlendirme türü olarak tercih etmiş olabilirler.

Öğrencilerin en az tercih etttikleri değerlendirme türü çoktan seçmeli madde türüdür. Öğrencilerin ilkokul seviyesinden üniversite düzeyine kadar girmiş olduğu geniş ölçekli sınavların neredeyse tümünün çoktan seçmeli sınav olmasına rağmen,

değerlendirme türü olarak en az tercih etmeleri, araştırmadan elde edilen önemli sonuçlardan biridir. Öğrencilerin en sondan ikinci olarak tercih ettikleri değerlendirme türü bireysel ev ödevidir. Öğrencilerin en sondan üçüncü sıraya yerleştirdikleri değerlendirme türü ise farklı soru türlerinin (çoktan seçmeli, kısa yanıtlı, doğru-yanlış, eşleştirme) aynı anda kullanılmasıyla gerçekleştirilen değerlendirmedir.

Bu araştırmanın sonuçlarına dayalı olarak yapılabilecek önerilerden biri, öğrencilerin birçoğunun belirtmiş olduğu gibi değerlendirme aşamasında tek tür soru formatı kullanılmasıdır. Araştırma bulgularında öğrencilerin, aynı anda birçok soru formatının kullanılmasını pek tercih etmedikleri belirlenmiştir. Bu nedenle, farklı soru formatları kullanıldığı durumlarda, bu soru formatları ile ilgili yönerge veya açıklamalara yer verilmesi, bu konudaki korku ve isteksizliği azaltacağı düşünülmektedir. Bu araştırma sadece eğitim fakültesinde öğrenim görmekte olan üniversite öğrencileri ile yürütülmüştür. Farklı fakülte veya yüksekokullarda öğrenim görmekte olan üniversite öğrencileri ile böyle bir çalışma yapılması alan yazına katkı sağlayabilir.

Anahtar Kavramlar: sıralama yargılarına dayalı ölçekleme, Değerlendirme tercihleri, Değerlendirme araçları