The Effect of Question Order on Exam Performance: A Six-Semester Study on Civil Engineering Students

Feride Maide MIZRAKCI[1] **Savaş BAYRAM***[2]

[1] Erciyes Üniversitesi Mühendislik Fakültesi İnşaat Mühendisliği, KAYSERİ
*[2] Erciyes Üniversitesi Mühendislik Fakültesi İnşaat Mühendisliği, KAYSERİ
*Sorumlu Yazar: Savaş BAYRAM

(Alınış / Received: 10.03.2025, Kabul / Accepted: 15.05.2025, Online Yayınlanma / Published Online: 31.07.2025)

Keywords
Civil Engineering,
Question Order,
Normality Test,
Homogeneity Test,
One-way ANOVA

Abstract: The term "cognition" refers to all types of knowing and consciousness. In cognitive engineering, information processing in a participant's or operator's mind or brain is often regarded to represent cognition. This study investigates whether presenting the same questions to students in different orders affects the exam performance of university students. The exam performances of the students to the multiple choice mid-term and final exam questions were evaluated. In this context, "Engineering Management and Human Relations", a final year course in the curriculum of the Civil Engineering Department of Erciyes University in Türkiye, was examined. Within the scope of the course, a total of 1,342 students covered six semesters between the 2016 and 2019 years. The questions were presented in the same sequence as the weekly flow of the course, (S) as A-type and randomly ordered (R) tests as B-type. The distributions of the answers given to the exam questions were analyzed via normality test and one-way ANOVA, and no significant differences between the test types were observed. This study covers only civil engineering students and whether similar results can be obtained for students in different disciplines should be examined in future studies.

Soru Sırasının Sınav Performansı Üzerindeki Etkisi: İnşaat Mühendisliği Öğrencileri Üzerinde Altı Yarıyıllık Bir Çalışma

Anahtar Kelimeler
İnşaat Mühendisliği,
Soru Sırası,
Normallik Testi,
Homojenlik Testi,
Tek yönlü ANOVA

Özet: "Biliş" terimi tüm bilme ve bilinç türlerini ifade eder. Bilişsel mühendislikte, bir katılımcının veya operatörün zihninde veya beyninde bilgi işlemenin genellikle bilişi temsil ettiği kabul edilir. Bu çalışma, öğrencilere aynı soruların farklı sıralarda sunulmasının üniversite öğrencilerinin sınav performansını etkileyip etkilemediğini araştırmaktadır. Öğrencilerin çoktan seçmeli ara sınav ve final sınav sorularına yönelik sınav performansları değerlendirilmiştir. Bu bağlamda, Türkiye'de Erciyes Üniversitesi İnşaat Mühendisliği Bölümü müfredatında son sınıf dersi olarak yer alan "Mühendislik Yönetimi ve İnsan İlişkileri" incelenmiştir. Ders kapsamında 2016 ve 2019 yılları arasında toplam 1342 öğrenci altı yarıyılı kapsamıştır. Sınav soruları, dersin haftalık akışıyla aynı sırada, A tipi olarak sıralı (S) ve B tipi olarak rastgele sıralı (R) testlerden oluşmuştur. Sınav sorularına verilen cevapların dağılımları normallik testi ve tek yönlü ANOVA ile analiz edilmiş ve test türleri arasında anlamlı bir farklılık gözlenmemiştir. Bu çalışma sadece inşaat mühendisliği öğrencilerini kapsamakta olup farklı disiplinlerdeki öğrenciler için benzer sonuçların elde edilip edilmeyeceği gelecek çalışmalarda incelenmelidir.

1. Introduction

The term "cognition" refers to all types of knowing and consciousness [1]. More widely, cognition refers to a person's knowledge, awareness, and opinions about an object, as well as that person's attitude toward the thing, or more specifically, the facts, knowledge, and beliefs about the object [2]. In a variety of contexts, including those found in the disciplines of linguistics, psychology, and education, the processes of cognition are examined from numerous angles. In cognitive engineering, information processing in a participant's or operator's mind or brain is often regarded to represent cognition [3].

Test anxiety in education is a problem frequently encountered by families, students, and educational institutions [4]. Studies show that there is a positive relationship between test anxiety and cognition [5]. The most commonly used exam type for large groups of students is multiple-choice tests [6]. When administering multiple-choice tests, instructors often use alternative formats without taking into account the potential impact of item order on test results [7]. Scientific research on organizing the order of exam items has a long history. Studies on this subject are presented below in chronological order.

According to MacNicol (1956), as a pioneering study, ordering the exam questions randomly and from easy to tough produced a considerably higher test score than ordering them from tough to easy [8]. Brenner (1964) on the other hand pointed out that no significant differences were observed in terms of random, easy to tough, tough to easy [9]. Moreover, Spies-Wood (1980) concluded that ordering the questions from easy to tough produced a considerably higher test score than randomly prepared questions [10]. Laffittee (1984), from a different perspective, ordered the questions in four different ways; tough to easy (in terms of section and in general separately) and random (in terms of section and in general separately), and concluded that no significant differences were observed [11]. Balch (1989) conducted a study examining the effect of item order on student performance in multiple-choice tests. In this context, he administered a multiple-choice test consisting of 14 units with a total of 75 questions to students with three different ordering methods. The approaches he used in ordering are as follows: (1) ordering the questions by the order in which they were covered in the course and the order in the textbook (sequential), (2) randomly mixing the questions in each unit and changing the order of the unit (section contiguity), and (3) completely random ordering of the questions. As a result, it was found that the order in which the questions were arranged by the course and book order resulted in significantly higher test scores compared to the others [7]. Neely, Springston, and McCann (1994) investigated the effects of sequenced (S) and random (R) format tests as well as the test anxiety levels of the students. It was concluded that no significant differences were observed for test formats. However, it was determined that there is a significant common effect between the ordering of the questions and anxiety levels [12]. In terms of current studies, Pettijohn II and Sacco (2007) investigated the effects of sequentially ordered, randomly ordered, and reverse-ordered test exams on students' exam performances, exam difficulty perceptions, and exam completion times. Significant differences were observed only for exam difficulty perceptions and the order of questions [13]. Tan (2009) investigated how question order affects item difficulty and item discrimination skills in multiple-choice tests. The sample was 996 first-year students from the Faculty of Education at Uludağ University. He found that question order in multiple-choice tests can have a statistically significant effect on both item difficulty and item discrimination. Significant differences were observed in 11 out of 50 questions in terms of item difficulty and only three questions in terms of item discrimination difficulty [6]. Şad (2020), in his study examining the effect of question order on test performance and item statistics, applied two different test forms in which the same multiple-choice questions were ordered from easy to difficult and from difficult to easy for 554 students. As a result, he found that item order did not have a significant effect on the achievement levels of test takers [14].

The aforementioned literature review indicates that there are no desirable studies on the effect of question order on item statistics. In the context of measurement and evaluation, the impact of structural features of tests on student performance is as important as the reliability and validity of tests. In this case, the order of test items is one of the variables that has been emphasized for a long time but has not been sufficiently analyzed. This study investigates whether there are significant differences in basic item statistics when multiple-choice test items are presented to students in different orders. The findings of the study will not only provide an understanding of the effect of question order on student performance but also provide recommendations for assessment and evaluation experts, professionals preparing question banks, and academics involved in test preparation. Furthermore, this study aims to make an original contribution by emphasizing whether the effect of question ordering should be considered one of the possible reasons for unexpected item statistics. In this study, some items obtained from classical test theory statistics, the effect of the order in which the questions were asked in the test to put it into practice. One of the main features of the performed studies is that they are handled within

the scope of a single exam. The present study, on the other hand, aims to examine both a 6-semester period and both the mid-term and final exams in the same semester.

2. Material and Methods

The undergraduate course, "Engineering Management and Human Relations", is offered as an elective course in both fall and spring semesters in the fourth and final year at the Civil Engineering Department of Engineering Faculty, Erciyes University. The purpose of the course is to inform students about project management, human resources management, and marketing management, ethics, business ethics besides the points to be considered during presentations. The weekly content of the course is provided in Table 1.

Table 1. Weekly Course Content for Engineering Management and Human Relations Course

Week	Topic
1	Project management
2	Project management
3	Morality, ethics, business ethics
4	Morality, ethics, business ethics
5	Entrepreneurship
6	Entrepreneurship
7	Marketing management
8	Marketing management
Mid-Term Exams	
9	Successful management techniques
10	Successful management techniques
11	Decisive and ordering in management
12	Public relations in management
13	The concept of presentation and presentation techniques
14	The concept of presentation and presentation techniques
Final Exams	

The research question that inspired this paper is “Does item order affect performance on multiple-choice exams?” Two test forms (Forms A and B), namely sequentially ordered (S) as A-type and randomly ordered (R) tests as B-type, were developed for this purpose. Please note that the questions in the same sequence as the weekly flow of the course and considered as (S). The same questions were included in both test forms. In both test forms A and B, the order of the questions (A, B, C, D or E) was not changed. Thus we focused on the mid-term and final exam results of the "Engineering Management and Human Relations" course for 6 semesters. The number of questions and student frequencies for each semester are presented in Table 2.

Table 2. Numerical Data on Examinations with Sequential and Randomised Test Forms

		Mid-Term Exam			Final Exam		
Nr.	Type	Number of Questions	Frequency (A-type)	Frequency (B-type)	Number of Questions	Frequency (A-type)	Frequency (B-type)
1	2016-2017 Fall	20	59	64	25	65	58
2	2016-2017 Spring	25	39	36	25	36	38
3	2017-2018 Fall	25	58	57	30	55	59
4	2017-2018 Spring	25	53	51	30	54	52
5	2018-2019 Fall	30	82	80	30	81	81
6	2018-2019 Spring	30	45	46	30	48	45
Sub-Total		**---**	**336**	**334**	**---**	**339**	**333**
TOTAL NUMBER OF STUDENTS		**1,342**					

In the exam, the students made their answers in the optical form, the exam papers were read with an optical reader and the data were transferred to the electronic environment. The distributions of the answers given to the exam questions were analyzed and it was evaluated whether there were significant differences between the correct answers given. The statistical analyses have been performed using IBM SPSS Statistics Version 15 software.

3. Results

This section presents the results of normality analysis and ANOVA analysis. First, the suitability of the data set for normal distribution was evaluated by calculating skewness and kurtosis values, and the results are presented in Table 3.

Table 3. Results of Normality Analysis

Nr.	Term	Type	Frequency	Skewness	Kurtosis
1	2016-2017 Fall	Mid-Term Exam	123	-0.387	-0.087
		Final Exam	123	-0.358	**-0.025**
2	2016-2017 Spring	Mid-Term Exam	75	-0.630	1.320
		Final Exam	74	-0.738	-0.347
3	2017-2018 Fall	Mid-Term Exam	115	-0.654	0.781
		Final Exam	114	-0.632	0.334
4	2017-2018 Spring	Mid-Term Exam	104	**-0.018**	-1.034
		Final Exam	106	-0.611	0.347
5	2018-2019 Fall	Mid-Term Exam	162	0.055	-0.343
		Final Exam	162	-0.466	0.170
6	2018-2019 Spring	Mid-Term Exam	91	-0.161	0.157
		Final Exam	93	-0.611	0.852

The previous studies indicate that the skewness/kurtosis values in the range of ±2 show a normal distribution [15, 16]. The skewness and kurtosis values were obtained for the current dataset as -0.738 to +0.055, and -1.034 to +1.320 respectively. These values confirm that the data 'fit' the normal distribution. The histograms of the datasets for the skewness value of -0.018 and kurtosis value of -0.025, which are closest to zero, are presented in Figure 1, respectively.

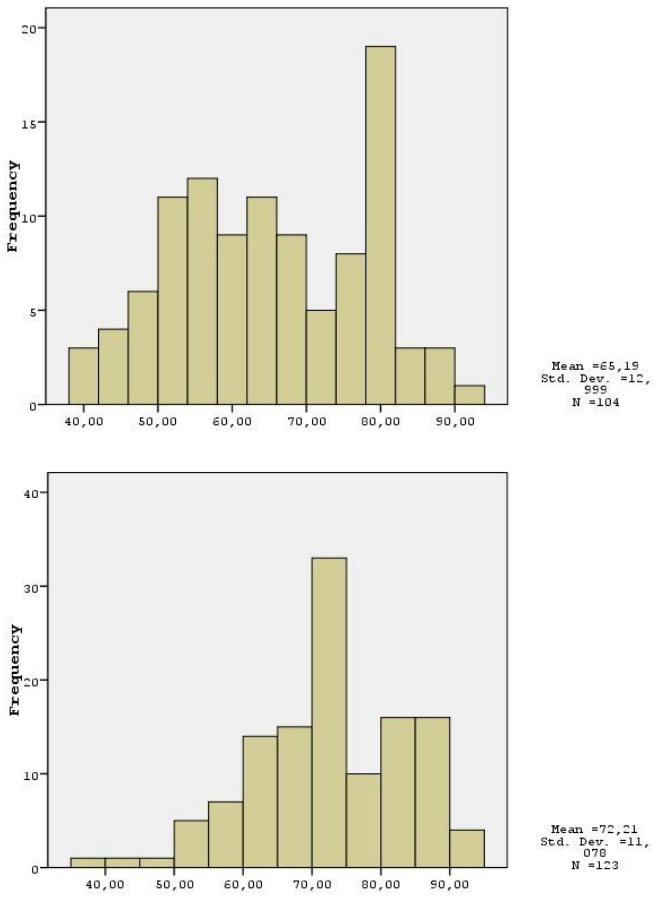

Figure 1. Histogram Distribution of A and B Group Exam Scores

The two categories of statistical tests are (i) Parametric tests (t-test, analysis of variance, Pearson correlation, ANOVA, etc.) and (ii) Non-parametric tests (Mann-Whitney U, Kruskal-Wallis H, Wilcoxon, Spearman correlation analysis, etc.), [17]. In this study, one-way analysis of variance (One-Way ANOVA), which is one of the parametric

tests, was selected within the scope of the analysis since the data distribution is suitable for normal distribution. In this context, the null hypothesis '*H_0: There is no significant difference between the correct answers given to Group A and Group B tests in the same exam*' was tested.

Table 4. Results of One-way ANOVA

Nr.	Term	Type	Homogeneity (sig.)	ANOVA (sig.)
1	2016-2017 Fall	1.1. Mid-Term Exam	0.883	0.204
		1.2. Final Exam	0.300	0.670
2	2016-2017 Spring	2.1. Mid-Term Exam	0.576	0.479
		2.2. Final Exam	0.730	0.592
3	2017-2018 Fall	3.1. Mid-Term Exam	**0.034**	0.319
		3.2. Final Exam	0.115	0.659
4	2017-2018 Spring	4.1. Mid-Term Exam	**0.038**	0.479
		4.2. Final Exam	**0.035**	0.798
5	2018-2019 Fall	5.1. Mid-Term Exam	0.147	0.490
		5.2. Final Exam	0.941	0.558
6	2018-2019 Spring	6.1. Mid-Term Exam	0.674	0.096
		6.2. Final Exam	0.434	0.500

A homogeneity test should be performed to decide whether ANOVA can be used [18]. According to Table 4, since the significance values for items other than items 3.1, 4.1, and 4.2 were greater than 0.05 (sig.>0.05), a homogeneous distribution was observed. Therefore, ANOVA significance values for the related items are important. According to Table 4, since the significance values for all items are greater than 0.05 (sig.>0.05), there is no significant difference between the scores of the students and the test type. In other words, hypothesis H_0 (No significant difference between the correct answers given to Group A and Group B tests in the same exam) is accepted.

4. Discussion and Conclusion

Classical test theory (CTT), one of the most widely used theories in measurement and evaluation, provides a basic framework for examining the reliability and validity of tests. CTT, which is widely used in education, psychology and other social sciences, is frequently applied to simplification and scale development [19, 20]. In this study, examining the effects of presenting exam questions in the same order as the weekly flow on students' exam performance, a field study was conducted covering 12 examinations over 6 semesters within the scope of the 'Engineering Management and Human Relations' course in the civil engineering curriculum. The suitability of the dataset for normal distribution was evaluated through skewness and kurtosis parameters, and it was observed that the entire dataset belonging to 12 semesters showed a normal distribution. Within the scope of the field research, examinations in the form of different types of tests were handled and it was investigated whether there were significant differences between the answers given by the students to the exams prepared randomly and according to the subject order in the course. The significance values of the related datasets were analyzed and it was concluded that there was no significant difference between the correct answers given to Group A and Group B tests in the same exam. In other words, it can be stated that question order does not affect student scores. There are different results in the literature on the effect of question order on student performance. The findings of this study, in line with Brenner (1964) and Laffitte (1984), suggest that question order does not affect student performance. However, Balch (1989) found that the presentation of questions in the order of the course and the book led to higher achievement levels than randomly ordered tests. Mean scores of 70.00% and 71.17% were obtained from sequentially ordered (S) A-type and randomly ordered (R) B-type exams respectively. Although this score difference is not statistically significant, it is thought that this difference may be pedagogically significant and cannot be ignored. In this case, the possibility that item order may make a difference, especially on test solution strategies and may have indirect effects on the student's cognitive process, should be taken into consideration. The student's motivation, attention level or use of prior knowledge about the subject may be affected by the test order. Although the close average scores obtained indicate that the effect of question ordering may be limited, more pronounced effects may be observed at the individual level or in certain subgroups (e.g. students with low achievement levels).

The findings of this study reveal that item order does not have a statistically significant effect on overall achievement level. However, in a field such as project management, which requires multidimensional thinking and decision-making skills and has a high cognitive intensity, the effects of test structures should not be limited to statistical significance. In particular, project management education requires students not only to acquire

knowledge but also to have the ability to structure this knowledge and apply it to different situations. In this context, the order of test questions may indirectly affect the way students perceive conceptual integrity and the way they associate information.

One of the most striking features of this study is that it differs significantly from previous studies in the literature in terms of sample size. For example, while Laffitte (1984) included 82 participants in his study, Pettijohn and Sacco (2007) included 66 participants in their study, and Şad (2020) included 554 participants in his study, the present study analyzed the data obtained from 1342 university students. This increases the generalizability of the findings and significantly increases the statistical power of the results. Therefore, this study has the potential to make a valuable contribution to the literature.

However, the study has some limitations. The study covers only civil engineering students and similar research can be conducted on larger sample groups, including students from different disciplines. Future studies can also examine the effects of individual differences such as student achievement level, learning style and cognitive skill levels on responses to question ordering. In terms of different test formats, open-ended questions, scenario-based assessments and project-based assignments can be used to examine the effect of question ordering. This method can help yield more meaningful data, especially in areas that require applied knowledge, such as project management.

Acknowledgment

The authors received no financial support for the research, authorship, and/or publication of this article.

References

[1] SI, A. S. A. (2022). *Cognitive Output of Nasopharyngeal Carcinoma Patients after Chemoradiation Therapy at Dr. Sardjito General*. Doctoral dissertation, Universitas Gadjah Mada, Yogyakarta, Indonesia.

[2] Liao, C. W., Liao, Y. H., Chen, B. S., Tseng, Y. J., & Ho, W. S. (2022). Elementary Teachers' Environmental Education Cognition and Attitude: A Case Study of the Second Largest City in Taiwan. *Sustainability*, 14(21), 14480.

[3] Blomberg, O. (2011). Conceptions of Cognition for Cognitive Engineering. *The International Journal of Aviation Psychology*, 21(1), 85-104.

[4] Kahramanoğlu, H. (2018). *The Relation Between Metacognition and Positive Psychological Capital with Test Anxiety.* Master's Thesis, Institute of Social Sciences, Haliç University, İstanbul, Türkiye. (in Turkish)

[5] Savaşan, İ. (2019). *The Relationship Between Negatİve Metacognition Level and Test Anxiety*, Master's Thesis, Institute of Social Sciences, Çağ University, Mersin, Türkiye. (in Turkish)

[6] Tan, Ş. (2009). The Effect of Question Order on Item Difficulty and Discrimination. *Education Sciences*, 4(2), 486-493.

[7] Balch, W. R. (1989). Item Order Affects Performance on Multiple-Choice Exams. *Teaching of Psychology*, 16(2), 75-77.

[8] MacNicol, K. (1956). Effects of Varying Order of item Difficulty in an Unspeeded Verbal Test. *Unpublished Manuscript, Educational Testing Service, Princeton, NJ*.

[9] Brenner, M. H. (1964). Test Difficulty, Reliability, and Discrimination as Functions of item Difficulty Order. *Journal of Applied Psychology*, 48(2), 98.

[10] Spies-Wood, E. (1980). Learned Helplessness and item Difficulty Ordering. *Psychologia Africana*, 19, 29–40.

[11] Laffitte Jr, R. G. (1984). Effects of Item Order on Achievement Test Scores and Students' Perception of Test Difficulty. *Teaching of Psychology*, 11(4), 212-214.

[12] Neely, D. L., Springston, F. J., & McCann, S. J. (1994). Does Item Order Affect Performance on Multiple-Choice Exams? *Teaching of Psychology*, 21(1), 44-45.

[13] Pettijohn II, T. F., & Sacco, M. F. (2007). Multiple-choice Exam Question Order Influences on Student Performance, Completion Time, and Perceptions. *Journal of Instructional Psychology*, 34(3), 142-149.

[14] Şad, S. N. (2020). Does Difficulty-Based Item Order Matter in Multiple-Choice Exams? (Empirical evidence from university students). *Studies in Educational Evaluation*, 64, 100812.

[15] Razali, N. M., & Wah, Y. B. (2011). Power Comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests. *Journal of Statistical Modeling and Analytics*, 2(1), 21-33.

[16] Mishra, P., Pandey, C. M., Singh, U., Gupta, A., Sahu, C., & Keshri, A. (2019). Descriptive Statistics and Normality Tests for Statistical Data. *Annals of Cardiac Anaesthesia*, 22(1), 67-72.

[17] Pallant, J. (2010). *SPSS Survival Manual: A Step By Step Guide to Data Analysis Using the SPSS Program.* 4th edition, McGraw Hill, New York.

[18] Aksoy, S. (2011). *Tests for Homogeneity of Variance Under Normality and a Comparison*. Master's Thesis, Gazi University, Ankara, Türkiye. (in Turkish)

[19] Baykul, Y. (2015). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: Pegem Akademi. (in Turkish)

[20] Wu, Y., Tang, J., Du, Z., Chen, K., Wang, F., Sun, X., & Wu, Y. (2025). Development of a Short Version of the Perceived Social Support Scale: Based on Classical Test Theory and Ant Colony Optimization. *BMC Public Health*, 25(1), 232.