SAKARYA UNIVERSITY JOURNAL OF EDUCATION

Original Research Doi: 10.19126/suje.1007897

Received: 10.10.2021 Accepted: 12.04.2022

Published: 15.04.2022 April 2022• 12(1) • 130-149

Investigation of Psychometric Properties of Multiple-Choice Items Developed By Turkish Teachers[*]

Kenan Burak YÜKSEL[**] Nuri DOĞAN[***]

Abstract. The main purpose of this research is to determine the psychometric properties of the items (questions) developed by Turkish teachers. The research was conducted with 320 teachers who teach Turkish to 8th grade students in public schools in Ankara in during the 2020-2021 academic year. Turkish subtest items of LGS (High School Enterance System) central exam verbal section were evaluated by field experts and a blueprint was created. Each achievement in the blueprint was randomly assigned to a teacher and he/she was asked to develop a multiple-choice item which has moderate difficulty and high discrimination to measure this achievement. The multiple-choice items written by the teachers were applied to the 8th grade students and the psychometric properties of the items were estimated from the students' responses. According to the results of the research, it was found that the content validity ratios, reliability and validity coefficients, item factor loadings, and differential item function values of the items developed by the teachers were at an insufficient level. Item difficulties were easy and the discrimination indexes were at a level that could be taken to the test by only correction.

Keywords: Item analysis, item psychometric properties, item writing, teacher-made tests.

[*] The ethics committee approval for this study was obtained from the Ethics Committee of the Rectorate of Hacettepe University, dated 10/09/2020 and numbered 35853172-300.
This study is produced on the Ph.D. Thesis titled "Effect Of Item Writer Characteristics On Item Psychometric Properties", prepared by Kenan Burak YÜKSEL under the supervision of Prof. Dr. Nuri DOĞAN in Hacettepe University, Institute of Education Sciences, Department of Measurement and Evaluation in Education.

[**] Orcid ID: https://orcid.org/0000-0003-4557-6663, Dr., Hacettepe University, Turkey, kbyuksel80@gmail.com

[***] Orcid ID: https://orcid.org/0000-0001-6274-2016, Prof. Dr., Hacettepe University, Turkey, nuridogan2004@gmail.com

Yüksel, K. B., & Doğan, N. (2022). Investigation of Psychometric Properties of Multiple-Choice Items Developed By Turkish Teachers. *Sakarya University Journal of Education, 12*(1), 130-149. doi: https://doi.org/10.19126/suje.1007897

1. INTRODUCTION

Evaluating the level of acquisition of the behaviors aimed at the end of the education process gives valuable feedback to the educational program, educators and students. Psychometry field fulfills this evaluation function by using measurement tools, in general terms tests. Because of the importance of their results, these tests are expected to be reliable, valid and fair (AERA, APA & NCME, 2014, p.11). However, a good number of the large-scale tests applied to monitor students' achievements or as a selection process for higher schools/education entrance or acceptance in Turkey, many items (questions) were canceled or correct answers (options) were changed even before the evaluation of how reliable, valid and fair measurements were made. Table 1 shows the number of items that were canceled or the declared correct answer changed in the exams applied for university placement and transition to secondary education between the years 2015-2019 in Turkey, which were announced to the public.

Table 1

Number of Items Canceled and Correct Answer Changed Between 2015-2019

Exam Name	Exam Year	Numbers of Canceled Item	Numbers of Items Correct Answer Changed
YKS	2019	2	
LYS	2017	2	
LYS	2016	2	2
YGS	2016	2	1
ALES	2016	1	
LYS	2015	3	
YGS	2015	1	
TEOG	2015	2	

It is seen in the table that many items were canceled or correct answers were changed in all the other years except 2018. Due to the cancellations/changes that damage the reliability of the measurement process, the time and efforts the respondents will waste both in the preparation period for the exam and in answering the item cannot even be rewarded. Similar losses can also be mentioned for test developers such as measuring centers (ÖSYM, MEB, etc.) and item writers responsible for development of the item and the test. In addition, this situation causes a negative perception towards central exams (Baş & Kıvılcım, 2019, p.655; Büyüköztürk, 2016, p.355; Karataş & Güleş, 2013, p.117). Furthermore, some studies on the items which were not cancelled or changed in large-scale tests also indicate that these items and tests are not at a sufficient level at all in terms of reliability and validity (Kaya, 2017, p.126; Özkan & Güvendir, 2014, p.41; Şata, 2016, p. 58; Yorgancı, 2015, p.54).

When the items and tests developed by teachers, whose one of the duties is to prepare students for these exams, are examined, similar results are obtained. Some studies on teachers' assessments and evaluation competencies indicate that they perceive themselves as inadequate in writing items (Çakan, 2004, p.108; Karamustafaoğlu, Çağlak & Meşeci, 2012, p.175) and that they need to be educated in item writing (Akçadağ, 2010, p.46; Anıl & Acar, 2010, p.58; Çelikkaya, Karakuş & Demirbaş, 2010, p.27). The 2023 Education Vision (MEB, 2018, p. 34) announced by the Ministry of National Education (MEB) on October 23, 2018 aims to reorganize all exams in the education system, in terms of purpose, content, and structure depending on question types and the benefit they will provide. Without knowing the level of "monitoring and evaluating learning and development", which is one of the qualifications of the teachers' (MEB, 2017, p.7), this methodology should not be implemented. For this reason, it is necessary to determine the level of success of teachers by examining their item writing skills. Studies so far have examined only one or a few features of teacher-made items (Aldım, 2010; Aybek, Yaşar & Kartal, 2021; Berberoğlu, 1996; Çağlar & Kılıç, 2019; Keskin, 2013; Şata, 2016; Öğretmen, 1995; Tokcan & Çevik, 2013). There is no research examining all psychometric properties of items as a whole in Turkey. Therefore, in this study, Turkish teachers were asked to write multiple-choice items. Then, these items were given to students, and psychometric properties were estimated from student responses. Thus, it has been tried to determine to what extent teacher-made items can measure student achievement.

Psychometry, which is formed by the combination of the Greek words soul (psyche (ψυχή)) and measurement (metron (μέτρον)), is defined as "measurement of psychological characteristics and skills" (Stuart-Hamilton, 2007, p.214). On the other hand, psychometrics is the field that deals with the qualities (type of information obtained or score, reliability and validity) of measurement tools used to measure human characteristics (Furr & Bacharach, 2013, p.7). AERA, APA and NCME (2014, p.11), on the other hand, mention three principles regarding the characteristics of measurement tools: a) validity, b) reliability/precision and error of measurement, c) fairness in testing. When the definitions and features are considered together, it can be mentioned that there are many statistics/parameters that can be cited as evidence for the psychometric properties that can be gathered under the headings of item scores and score distributions, reliability, validity, and fairness. Item difficulty, discrimination and distribution of answers to options can be listed as mandatory indicators regarding the functioning of an item (Doğan & Tezbaşaran, 2003, p.58). It can be argued that these indicators are important but not sufficient. For example, these indicators cannot reveal whether the item provides an advantage/disadvantage to the groups that differ in terms of a feature different from the measured feature or how the score distribution is. So, revealing statistics that provide information about the different features of the item (the functionality of distractors, the distribution of item scores, etc.) will contribute to more accurate decisions, especially inclusion and exclusion of the item in the test. In this context, with the scores obtained as a result of the application of multiple-choice items, psychometric properties to be presented as evidence for score distributions, item statistics, reliability, validity and fairness were determined by the researcher by evaluating all the techniques in the field, and the definitions, calculation methods, criteria and interpretations of these properties are presented in Appendix 1.

2. METHOD

Universe and Study Group

The universe of the research consists of 2725 (1937 female, 788 male) Turkish teachers and 66,517 (32317 female, 34200 male) 8th grade students who teach and learn in Ankara public schools in the 2020-2021 academic year. The study group was formed with 320 Turkish teachers and 4142 students who participated voluntarily from the universe. The demographic and graduation information of the study group are presented in Table 2. The ethics committee approval for this study was obtained from the Ethics Committee of the Rectorate of Hacettepe University, dated 10/09/2020 and numbered 35853172-300.

Table 2

Demographic and Graduation Information of the Study Group

Exam Name	Teachers		Students	
	n	%	n	%
Gender				
Female	173	54	2005	48
Male	147	46	2137	52
Age				
25-30	13	4		
31-40	69	22		
41-50	171	53		
51-60	65	20		
60+	2	1		
Graduation				
Associate degree	14	4		
Bachelor's degree	261	82		
Master’s degree	45	14		
Graduated Field				
Department of Turkish teacher	179	56		
Turkish language and literature	51	16		
Department of Turkish language and literature	83	26		
Other*	7	2		

** Contemporary Turkish Dialects, German*

Data Collection and Analysis

Before the data collection process, Ethics Approval was obtained by applying to the Hacettepe University Ethics Committee to check the compliance of the research with the ethical rules. Research Permission was obtained from the Ministry of National Education for teachers and students to participate in the research. Due to the Covid-19 epidemic, data were collected remotely instead of face to face engagement. Data collection tools were put on the https://www.classmarker.com/web link and data were collected between 15 October 2020 and 15 July 2021 by directing to their e-mail addresses of respondents. An expert group was formed by inviting 12 Turkish field experts [3 Ph.D. (Classic Turkish literature, Turkish language, contemporary Turkish dialects), 9 Master's graduates (Turkish folk literature (3), Classic Turkish literature (2), contemporary Turkish dialects (2), Turkish language and literature (2))] to make item-achievement matching in order to create the blueprint and to evaluate the items in order to determine the content validity ratios of the items developed by the teachers.

It would be appropriate to design a form (test) that could measure equivalent to 2021 LGS in terms of scope for calculating item validity indexes. Since the 2020 LGS was prepared only on the basis of the 1st semester curriculum, it was decided that teachers should develop items within the scope of the 2019 LGS verbal section Turkish subtest. The arithmetic mean of the Turkish subtest scores calculated on the answers given by 1.029.555 students to the test administered on 1 June 2019 was 11.75; its standard deviation is 5.15; average difficulty 0.59; average discrimination power 0.59; The KR-20 internal consistency coefficient was estimated as 0.87 (MEB, 2019a, p.15). In order to determine which achievements the 2019 LGS verbal section Turkish subtest items were developed to measure, they were asked to match the items with 86 achievements determined in the curriculum (MEB, 2019b, p.47) to 12 field experts. The highest of the matching results made by the experts was 100%, and the lowest was 67%. In order to explore the level of agreement between the field experts' matches, both Krippendorff's Alpha and the Fleiss Kappa coefficients were calculated since the data obtained was categorical (discrete) and the number of raters is more than two. Coefficient of Krippendorff's Alpha estimated 0.883 and Fleiss Kappa calculated 0.876. These findings indicate that there is an almost perfect agreement between expert opinions for item-achievement matching (Krippendorff, 2004, p.241; Landis, & Koch, 1977, p.165). By combining item-achievement matches, the blueprint of the tests to be developed in this study was obtained. Teachers were given a random achievement, and asked to develop an item that has moderate difficult and high discrimination. 16 forms were created with 320 items written by the teachers and applied to the students. A one-way analysis of variance (one-way ANOVA) was conducted for unrelated samples in order to determine whether the scores obtained by the students differed according to the forms they answered. According to the results of the analysis, no significant difference was found in terms of the applied form ($F_{14,4112}$=0.868; $p>0.05$). According to this result, it can be said that the psychometric properties of the items estimated from the different form groups will be independent of the success difference of the groups.

While estimating the psychometric properties of the items, the calculation methods in the Appendix were taken as basis. TAP 16.11.13 software for calculating difficulty and discrimination; Microsoft Office 2016 Excel for calculating variance, skewness, kurtosis, reliability, validity and distraction indexes; Factor 10.4 programs were used to calculate item factor loads based on the tetrachoric correlation matrix. To check whether the data resulted from each twenty items of sixteen forms is suitable for factor analysis, KMO and Bartlett sphericity tests were used. KMO values were found to be mediocre with 0.62-0.69 (Kaiser, 1974), and the Bartlett sphericity test results were found to be significant at the level of α=0.01 (Bartlett, 1950). Factor eigenvalues of first factors were found to be 3.47 and 4.33, explained variance ratios were 20.75 % and 24.12 %. Unrotated item loadings on the single (first) factor were interpreted to determine relationship between tests and the items of these tests. In the calculation of the item validity indexes, LGS (High School Enterance Exam), which was applied to the students on June 6, 2021 and the results of which were announced on June 30, 2021, verbal section Turkish subtest scores were taken as an external criterion. The arithmetic mean of the Turkish subtest scores calculated on the answers given by 1.038.492 students to the test was 14.86; its standard deviation is 3.88; average difficulty 0.47; average discrimination 0.41; The KR-20 internal consistency coefficient was estimated as 0.82 (MEB, 2021, p.13).

While calculating the CVR, it was taken as a basis how many experts selected the "necessary" option for each item. In the research conducted by Ayre & Scally (2014, p.85), a new critical value table was created by eliminating the distributional error of the critical values determined by Lawshe (1975, p.568) in the opinion of more than 10 experts. The EASY-DIF program (González et al., 2011, p.1) was used to calculate the Differential Item Function (DIF) values.

3. FINDINGS

Scoring the answers given by the students to the 320 items developed by the teachers and the descriptive statistics of the predicted item psychometric properties are presented in Table 3 as item statistics, reliability, validity, impartiality and score distributions, respectively.

Table 3

Descriptive Statistics of Item Psychometric Properties

Item Psychometric	Min	Max	Averag	SD	Skewne	Kurtosis
Item Difficulty Index	0,27	0,97	0,70	0,132	-0,179	-0,211
Item Discrimination	0,08	0,36	0,219	0,051	-0,008	-0,238
Item Distraction Index	0,18	0,96	0,61	0,161	-0,184	-0,445
Item Reliability Index	-0,26	0,21	-0,003	0,072	-0,148	0,201
Item Factor Loading	-0,42	0,42	0,006	0,131	0,069	0,198
Item Validity Index	-0,20	0,20	-0,001	0,074	0,079	-0,054
Content Validty Ratio	-0,50	1,00	0,57	0,438	-0,649	-0,713
Differential Item	-3,60	3,17	0,004	1,075	-0,047	0,164
Item Variance	0,12	0,25	0,223	0,021	-1,287	2,270
Item Skewness	-1,76	2,14	-0,038	0,721	0,118	-0,994
Item Kurtosis	-2,00	2,56	-1,480	0,516	2,690	13,466

It is seen that the item difficulty indexes take values between 0.27 (difficult) and 0.97 (very easy) and their arithmetic mean is at the easy level with a value of 0.70. Since the teachers were asked to develop items of moderate difficulty (0.40-0.60) in the study, it can be said that according to the results of the analysis, the item difficulties were not adjusted by the teachers in the study group in general.

It is seen that the item discrimination indexes range from 0.08 to 0.36, and their arithmetic averages are at the level of 0.219. Considering the average value, it can be said that the items developed by the teachers are not at a sufficient level in terms of discriminating the students who have an achievement from those who do not, and that the items are generally at a level that can be corrected and taken to the test (Crocker, & Algina, 2008, p.313). This finding coincides with studies with findings in Turkey where the discrimination of teacher-made test items is not at the expected level (Keskin, 2013, p.71; Şata, 2016, p.58).

When the distraction indexes are examined, it is seen that they have values between 0.18 and 0.96, and the arithmetic mean is at the level of 0.61. Since the average of acceptable distraction index is 0.44, it can be argued that the teachers in the study group developed items with balanced functioning distractors. This seems to be in line with the results of Coşkun (2021)'s research on the distractors of the LGS Turkish subtest items, with findings that none of the students chose the distractor for 4 items in the 20-item subtest, and the distractors of the other items showed an unbalanced distribution.

When the item reliability indexes in Table 3 are examined, it is seen that values are between -0.26 and 0.21 and the averages are (-0.003) almost 0. This finding indicates that the indexes of nearly half of the items have negative values, that is, they have an inverse relationship with the test scores they are in and measure another achievement. It is seen that a similar situation is valid for item factor loadings. According to unrotated item factor loadings, almost half of them have negative value. Since the factor loadings show how much a factor explains a variable, the test scores consisting of the sum of the item scores cannot explain almost half of the items or it can be said that they explain a different structure or achievement in other words.

It can be said that item reliability coefficients and the item factor loads are in a way that damages the reliability and construct validity of the tests respectively in which the items are included. The fact that the factor loadings were found in the opposite direction, close to half of the items, suggests that the students had difficulty in understanding the meaning expressed by the item or that they understood it differently. It can be said that the average of the validity indexes indicating the level of correlation with another test score is almost 0, the relationship with another test score, which is not included in about half of the items, changes in the opposite direction, and almost half of the items are insufficient to provide proof of validity. It seems parallel with the findings of the studies that teachers are insufficient to evaluate their students' performances reliably and validly (Mertler, 1999; Stiggins, 1999). Also, this finding supports Goodrich's (1977, p.69) claim that teachers are inadequate in developing reliable and valid items because they do not have as much training, equipment and time as expert item writers.

When Table 3 is examined, it is seen that the Content Validity Ratio (CVR) values are between -0.50 and 1, and the arithmetic mean is 0.57. It can be said that teachers are not at a sufficient level to develop items that can represent the content, since the minimum value that the CVR values calculated based on the opinions of 12 field experts should be at α=0.05 significance level is 0.667 (Ayre, & Scally, 2014, p.85). This finding indicates that the content validity of the items developed by teachers is not sufficient in Turkey (Aldım, 2010, p.60; Çağlar and Kılıç, 2019, p.1301; Keskin, 2013, p.71; Tokcan and Çevik, 2013, p.369) is consistent with the results.

Considering the differential item function values, 45 items (Appendix 2) with an absolute value greater than 1.5 were found, and 32 of them were found to work in favor of male students. This finding is in parallel with the studies with findings in which the items in the Turkish field exhibit differential item functions in favor of men (Berberoğlu, 1996, p.369; Öğretmen, 1995, p.58). However, Aybek, Yaşar and Kartal (2021, p.297) examined the DIF status of teacher-made items, which differed from the finding that no items that could cause DIF were found.

Item variance values are between 0.12 and 0.25 and the arithmetic mean is 0.22. Considering that the item variance can take the value of 0.25 maximum, it can be said that the items developed by the teachers are at a good level in this respect, as the variance increases, the power to reveal the differences between the respondents in terms of the quality measured by the item. However, the differences should not be considered in the sense that the item reveals the correct way with the knowledge levels of the students in terms of the achievement measured by the item. Considering the arithmetic means of the results of the skewness and kurtosis coefficients, which are the descriptors of the item score distributions, it can be said that the items are very lightly skewed to the left, in other words, the items developed in general are easy. It is seen that the skewness and kurtosis are in parallel with the results of the item difficulty indexes.

4. RESULTS, DISCUSSIONS AND SUGGESTIONS

In this study, it was tried to determine the level of teachers' item writing skills. For this reason, a multiple-choice item measuring an achievement was written by teachers. The calculated psychometric properties are discussed below and suggestions are made.

Although the teachers were asked to write items of moderate difficulty, it was determined that the item difficulty index, which is one of the item statistics, was at an easy level. In other words, teachers were found to be insufficient in adjusting the item difficulties. This may have arisen for two reasons. First, teachers may not be able to master students' knowledge levels. Secondly, even if they are proficient in knowledge level of students, they may be aware of how to write a medium difficulty item. Item discrimination levels, the second of the item statistics, were found to be insufficient. It can be said that the teachers are not at the expected level in terms of developing the item that will distinguish the students who know and who do not know the achievement. As with the item difficulty, this may be because the teachers did not know the students well enough. The reason for why the teachers did not get to know their students enough may be because the classes were held with remote access in the 2020-2021 Academic Year. It is important for teachers to be familiar with the level of students' understanding of the lessons taught and their level of success. This is also one of the teachers' competencies. For this reason, it can be suggested that while teaching, teachers should try to recognize what students know while teaching and with by using in-class assessments. It is expected that teachers may be able to assess which subjects of the lesson students have difficulty in understanding and what they misunderstand as a result of based on the interaction they have with the students. It is hoped that teachers will not be content with just that but will additionally identify the gaps at each student level. They can identify these gaps with verbal question-answer activities, or they can achieve this by asking students to explain or write what they understood about the topic. However, primarily it can be suggested that teachers should determine some criteria in their minds and compare the answers of the students with these criteria and hence evaluate their situation. In addition, it can be recommended to collect reliable and valid information about students by making item and test analysis of teacher-made tests. There are a few rules that can be suggested in order to develop more reliable and valid items. To begin with, "provide clear instructions for each item" can be the first suggestion. Expressing the questions to be asked to students, whether oral or written, as clear, simple and concise as possible will make it easier for them to grasp what exactly the question mean. A second recommendation can be "write items or ask questions that capture the essence of the topic being taught". In this way, it can be avoided to (overload/overcharge) students' cognitive skills with unnecessary details or unimportant parts of the subject. By following this approach, main efforts will focus to comprehend the core of the subject. Students at the later stages of their educational journey, can be provided with more granularity, if it is in line with the methodology of the teacher or/and the educational system. Last but not least suggestion can be "investigate the clarity and thoroughness of the item by asking feedback from students and colleagues". Asking students to describe what they understand from the item can also be one way. Thus, the difference between what they understand and what the teacher tries to emphasize can be revealed. Such kind of an approach can also be followed with the participation of other teachers, and by asking their feedback, the mistakes which are being made and the deficiencies in the respective field or/and the gap between what was expressed to the students and what they understood can be detected.

As the last item statistic, item distraction indexes were found to be sufficient. This indicates that teachers know the mistakes that students often make or are familiar with misunderstandings.

Almost half of the item reliability indices determined in the study as proof of reliability were calculated as negative. This means that the relationships between the items and half of the other items on the same test are in the opposite direction. It can be said that this inverse correlation damage reliability, and the results will vary from measurement to measurement. It can be suggested that teachers try to develop items that will accurately measure the achievement. If they can achieve this, the items in the test will be related to each other and reliability will increase.

Item factor loads calculated as evidence for construct validity were found to be at a similar level with item reliability coefficients. This result indicates that the students could not understand the structure or feature to be measured with the item or they understood it in a different sense. While writing the root of the item, the teachers should prepare it in accordance with the level of the students in terms of grammar and expression. Otherwise, students will not be able to answer correctly even if they know the feature measured by the item by understanding it differently. Half of the item validity coefficients with proof of concurrent validity are insufficient. Insufficient correlation with a reliable and valid test score indicates that the item was not designed well. Teachers need to be trained in the rules of article writing. Otherwise, it seems inevitable that they will develop items that will damage the validity of the item. Content validity ratios also tell us that the items are not at a level to represent the content. Teachers are either not familiar with the concept of achievement or, even if they know the phenomenon of achievement, they cannot develop an item to correspond to it. Content validity can simply be defined as how well the item represents the taught content. For this reason, teachers first need to think about how the subject of the lesson will be expressed. For example "the correct use of punctuation in a sentence" or "determining a title for an essay". Thus, the subject covered in the lesson can be expressed accurately and without confusing it with another subject. Then, a question sentence should be formed to meet this expression as "in which sentence is the dot used incorrectly?" or "which is one of the situations where a semicolon can be used in given blanks?”. In order to be sure that the question statement covers the subject, she or he should ask herself/himself the question "How can I create an item with this question sentence to measure another subject other than aimed". If teacher can develop an item that can measure another subject with this question, she or he should try to reconstruct the question sentence. This loop should be continued until a question sentence is formed in which sense no other subject can be tested anymore. At the end of this process, the item can be designed by obtaining a question sentence that only examines the subject stated at the beginning. In this way, teacher may increase the probability of ensuring that the item correctly covers the scope. For a better validity of content, it may be recommended to provide the teachers with hands-on item writing training during seminars. In addition, it can be suggested that the Ministry of National Education may prepare booklets which contain sample items that can accurately measure content for learning outcomes to guide teachers.

When the Differential Item Functioning values are examined, it is found that they are at a level provides an advantage to male students. It is expected from the item to measure the knowledge of the students without taking into consideration of another respondent feature. However, the research findings indicate that there are items that can measure in favor of a group of students. One reason for this can be that the items developed based on the situation are designed in a way that male students can comprehend better. Items may have been developed on a specific situation such as "football, computer game, etc." that female students are not interested in or do not have as much knowledge as boys do. While applying the items in the classroom, it can be suggested that the teachers examine the students' answers and examine whether they are differentiated according to the subgroup characteristics.

Since the distribution of item scores is a function of the item difficulty indices, it was found that they were at a similar level and that the items were of easy difficulty. It can be suggested that teachers should know how the item score distributions will give clues about the items and use them by calculating. Although it is desirable that the variance values are high, when evaluated together with other psychometric properties, it can be said that it would be appropriate for this differentiation to be in terms of the measured feature.

As a result, it can be said that teacher-made items are not at a good level except for distraction and variance, and measurements made with these items will give misleading results. They can write better items with the experience to be gained as a result of item writing training and following the success of the students in the classroom. Teachers need to develop the items at an appropriate level so that they can present the educational status of the students justly.

This study was carried out on the basis of classical test theory and in the field of Turkish. It can be suggested that a similar study may be carried out with item statistics estimated with item response theory or in different areas such as mathematics.

References

AERA, APA, & NCME. (2014). *Standards for educational and psychological testing.* Washington, DC, American Educational Research Association.

Akçadağ, T. (2010). Öğretmenlerin İlköğretim Programındaki Yöntem, Teknik, Ölçme ve Değerlendirme Konularına İlişkin Eğitim İhtiyacı [Teachers' Training Needs on Methods, Techniques, Measurement and Evaluation in Primary Education Curriculum]. *Bilig, 53*, 29-50.

Aldım, Ü. F. (2010). *İlköğretim 7. Sınıflarında Uygulanan SBS (Seviye Belirleme Sınavı) İngilizce Sorularının Bazı Değişkenlere Göre İncelenmesi [Examination of SBS (Placement Exam) English Questions Given in 7th Grades of Primary Education According to Some Variables]*. (Unpublished master's thesis). Fırat University, Elazığ.

Anıl, D., & Acar, M. (2010). Sınıf Öğretmenlerinin Ölçme Değerlendirme Sürecinde Karşılaştıkları Sorunlara İlişkin Görüşleri [Opinions of Classroom Teachers on the Problems Encountered in the Assessment and Evaluation Process]. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi, 5*(2), 44-61.

Aybek, E. C., Yaşar, M., & Kartal, S. (2021). Öğretmen Yapımı Bir Testteki Maddelerin Değişen Madde Fonksiyonu Bağlamında İncelenmesi [Examining the Items in a Teacher-Made Test in the Context of Changing Item Function]. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, (52)*, 281-300.

Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development, 47*(1), 79-86.

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Mathematical and Statistical Psychology, 3*(2), 77-85.

Baş, G., & Kıvılcım, Z. S. (2019). Türkiye'de Öğrencilerin Merkezi Sistem Sınavları ile İlgili Algıları: Bir Metafor Analizi Çalışması [Students' Perceptions of Central System Exams in Turkey: A Metaphor Analysis Study]. *Eğitimde Nitel Araştırmalar Dergisi, 7*(2), 639-667.

Baykul, Y. (2015). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması [Measurement in Education and Psychology: Classical Test Theory and Practice]*. (3rd edition). Ankara, Pegem Akademi.

Berberoğlu, G. (1996). The University Entrance Examinations in Turkey. *Studies in Educational Evaluation, 22*, 4, 363-373.

Büyüköztürk, Ş. (2016). Sınavlar üzerine düşünceler. *Kalem Eğitim ve İnsan Bilimleri Dergisi, 6*(2), 345-356.

Coşkun, S. (2021). *Liselere Geçiş Sisteminde İlçelerin Gelişmişlik Düzeyinin ve Cinsiyetin Değişen Çeldirici Fonksiyonuna Etkisi [The Effect of the Development Level of the Districts and the Gender on the Changing Distractor Function in the Transition System to High Schools]*. (Unpublished master thesis). Hacettepe University, Ankara.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning, 5191 Natorp Boulevard Mason, Ohio.

Çağlar, M., & Kılıç, A. (2019). Merkezi sınav ve öğretmen yapımı sınavların bazı değişkenler açısından incelenmesi: ortaöğretime geçiş sınavı örneği [Examining the central exam and teacher-made exams in terms of some variables: an example of secondary education entrance exam]. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 19*(4), 1288-1305.

Çakan, M. (2004). Öğretmenlerin Ölçme-değerlendirme Uygulamaları ve Yeterlik Düzeyleri: İlk ve Ortaöğretim [Teachers' Assessment and Evaluation Practices and their Competence Levels: Primary and Secondary Education]. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi, 37*(2), 99-114.

Çelikkaya, T., Karakuş, U., & Demirbaş, Ç. Ö. (2010). Sosyal Bilgiler Öğretmenlerinin Ölçme-değerlendirme Araçlarını Kullanma Düzeyleri ve Karşılaştıkları Sorunlar [The Levels of Social Studies Teachers' Use of Assessment-Evaluation Tools and the Problems They Encounter]. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi, 11*, 57-76.

Doğan, N., & Tezbaşaran, A. A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklemler bağlamında karşılaştırılması [Comparison of classical test theory and latent trait theory in the context of samples]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 25*, 58-67.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurements* (5th edition). Prentice-Hall of India.

Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: an introduction*. Sage Publications, Inc. 2455 Teller Road Thousand Oaks, California.

González, A., Padilla, J. L., Hidalgo, M. D., Gómez-Benito, J., & Benítez, I. (2011). EASY-DIF: Software for analyzing differential item functioning using the Mantel-Haenszel and standardization procedures. *Applied Psychological Measurement, 35*(6), 483.

Goodrich, H. C. (1977). Distractor Efficiency in Foreign Language Testing. *TESOL Quarterly, 11*(1), 69.

Gulliksen, H. (1950). *Theory of mental tests*. John Wiley&Sons, Inc., New York.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31-36.

Karamustafaoğlu, S., Çağlak, A., & Meşeci, B. (2012). Alternatif Ölçme Değerlendirme Araçlarına İlişkin Sınıf Öğretmenlerinin Öz Yeterlilikleri [Self-Efficacy of Classroom Teachers Regarding Alternative Assessment and Evaluation Tools]. *Amasya Üniversitesi Eğitim Fakültesi Dergisi, 1*(2), 167-179.

Karataş, S., & Güleş, H. (2013). Öğretmen atamalarında esas alınan merkezi sınavın (KPSS) öğretmen adaylarının görüşlerine göre değerlendirilmesi [Evaluation of the central exam (KPSS), which is used as a basis for teacher appointments, according to the opinions of teacher candidates.]. *Kuramsal Eğitimbilim Dergisi, 6*(1), 102-119.

Kaya, İ. (2017). 2010-2013 Yıllarında Sorulan YGS ve LYS Tarih Sorularının Ortaöğretim Tarih Dersi Öğretim Programları Açısından Değerlendirilmesi [Evaluation of YGS and LYS History Questions Asked in 2010-2013 in terms of Secondary Education History Course Curriculum]. *Journal of Analytic Divinity, 1*(1), 101-128.

Keskin, H. (2013). *İlköğretim ikinci kademe matematik öğretmenlerinin uyguladıkları sınavların psikometrik niteliklerinin incelenmesi [Examination of the psychometric qualities of the exams administered by the secondary school mathematics teachers.].* (unpublished master thesis). Akdeniz University, Antalya.

Khine, M. S. (2013). *Application of structural equation modeling in educational research and practice*. Rotterdam / Boston / Taipei, Sense Publishers.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, California, Sage.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174.

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575.

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease*. Journal of the national cancer institute, 22*(4), 719-748.

Mertler, C. A. (1999) Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education, 120*, 285–296.

MEB (Millî Eğitim Bakanlığı) (2017). *Öğretmenlik Mesleği Genel Yeterlikleri [General Competencies of the Teaching Profession]*. Ankara. Retrieved from http://oygm.meb.gov.tr/meb_iys_dosyalar/2017_12/11115355_YYRETMENLYK_MESLEYY_GENEL_YETERLYKLERY.pdf

MEB (Millî Eğitim Bakanlığı) (2018). *Güçlü Yarınlar için 2023 Eğitim Vizyonu [2023 Education Vision for Great Future].* Retrieved from http://2023vizyonu.meb.gov.tr/doc/2023_EGITIM_VIZYONU.pdf

MEB (Millî Eğitim Bakanlığı) (2019a). *2019 Ortaöğretim Kurumlarına İlişkin Merkezi Sınav. Eğitim Analiz ve Değerlendirme Raporları Serisi [2019 Central Exam, Education Analysis and Evaluation Reports Series for Secondary Education Institutions].* Retrieved from https://www.meb.gov.tr/meb_iys_dosyalar/2019_06/24094730_2019_Ortaogretim_Kurumlarina_Iliskin_Merkezi_Sinav.pdf

MEB (Millî Eğitim Bakanlığı) (2019b). *Türkçe dersi (1-8. sınıflar) öğretim program [Turkish Language lesson (1-8th grades) curriculum]. Ankara, Talim ve Terbiye Kurulu Başkanlığı.* Retrieved from https://mufredat.meb.gov.tr/Dosyalar/20195716392253-02-T%C3%BCrk%C3%A7e%20%C3%96%C4%9Fretim%20Program%C4%B1%202019.pdf

MEB (Millî Eğitim Bakanlığı) (2021). *Liselere Geçiş Sistemi (LGS) Merkezî Sınavla Yerleşen Öğrenci Performansı [High School Entrance System (LGS) Performance of Students Placed by Central Examination].* Retrieved from https://cdn.eba.gov.tr/icerik/2021/07/rapor/No_17-LGS_2021-merkezi_yerlestirme_211730.pdf

Öğretmen, T. (1995). *Differential Item Functioning Analysis of the Verbal Ability Section of the First Stage of the University Entrance Examination in Turkey*. (unpublished master's thesis). Middle East Technical University, Ankara.

Özkan, Y. Ş., & Güvendir, M. A. (2014). Türkiye'de Uygulanan Geniş Ölçekli Testlerin Çok Boyutluluğunun Analizi [Analysis of the Multidimensionality of Large-Scale Tests Implemented in Turkey]. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi, 29*, 31-47.

Ray, W. S., Hundleby, J. D., & Goldstein, D. A. (1962). Test skewness and kurtosis as functions of item parameters. *Psychometrika, 27*(1), 39-47.

Stiggins, R. J. (1999) Are you assessment literate? *High School Magazine, 6*(5), 20–30.

Stuart-Hamilton, I. (2007). *Dictionary of psychological testing, assessment and treatment*. (2nd Ed.) Jessica Kingsley Publishers.

Şata, M. (2016). Türk Eğitim Sistemi'nde Sınıf İçi ile Geniş Ölçekli Ölçme ve Değerlendirmeye Genel Bir Bakış [An Overview of In-Class and Large-Scale Measurement and Evaluation in the Turkish Education System]. *Current Research in Education, 2*(1), 53-60.

Tokcan, H., & Çevik, E. (2013). İlköğretim II. kademe sosyal bilgiler dersi öğretmenlerinin yazılı sınav sorularının programa uygunluğunun incelenmesi [Examination of the compatibility of the written exam questions of primary school second level social studies teachers to the program.]. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi, 32*(1), 58-91.

Thompson, B. (1981). *Factor stability as a function of item variance*. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas.

Widaman, K. F. (1993). Common Factor-Analysis versus Principal Component Analysis – Differential Bias in Representing Model Parameters. *Multivariate Behavioral Research, 28*(3), 263-311.

Yorgancı, O. K. (2015). *Sekizinci sınıf türkçe dersi ortak sınavı sorularının öğretim programına göre değerlendirilmesi [Evaluation of eighth grade Turkish lesson common exam questions according to the curriculum].* (unpublished master thesis). Gazi University, Ankara.

Zieky, M. (1993). *Practical questions in the use of DIF statistics in test development*. In P. W. Holland & H. Wainer (Eds.), Differential item functioning (pp. 337–347). Lawrence Erlbaum Associates, Inc.

Zumbo, B. D. (1999). *A handbook on the theory and methods of differential item functioning (DIF)*. Ottawa, National Defense Headquarters.

Appendix.1

Psychometric Property	Definition	Estimation Method	Criterion
Item Variance	It is a statistic that reveals the differences of item scores from the mean, or more broadly, the differences between individuals in terms of measured quality (Baykul, 2015, p.226; Thompson, 1981).	Item Variance (s_j^2) $s_j^2 = p_i.q_i$ *$q_i : 1\text{-}p_i$* (Baykul, 2015, p.225; Thompson, 1981)	$0,25 \geq s_j^2 \geq 0,00$ *It can be said that as the item variance increases, the power of revealing the differences between the respondents in terms of the quality measured by the item increases and decreases as it decreases. (Baykul, 2015, p.226; Thompson, 1981).*
Item Skewness Index	It is an indicator of whether the distribution of item scores is symmetrical. The fact that the distribution of item scores is symmetrical is important in test theories, especially when there is a normal distribution assumption (Baykul, 2015, p.228; Ray, Hundleby & Goldstein, 1962).	Item Skewness Index (α_{3j}) $\alpha_{3j}=\frac{1\text{-}2.p_{jx}}{\sqrt{p_{jx}.(1\text{-}p_{jx})}}$ *p_{jx} : item difficulty index* (Baykul, 2015, p.228; Ray, Hundleby & Goldstein, 1962)	$+\infty \geq \alpha_{3j} \geq -\infty$ *Distribution of item scores; if the coefficient is 0, it is symmetrical; as it gets closer to +/-∞, it moves away from symmetry (Baykul, 2015, p.228; Ray, Hundleby & Goldstein, 1962).*
Item Kurtosis Index	It is an indicator of whether the distribution of item scores is symmetrical. The fact that the distribution of item scores is symmetrical is important in test theories, especially when there is a normal distribution assumption (Baykul, 2015, p.229; Ray, Hundleby & Goldstein, 1962).	Item Kurtosis Index (α_{4j}) $\alpha_{4j}=\frac{1\text{-}6.p_{jx}.(1\text{-}p_{jx})}{p_{jx}.(1\text{-}p_{jx})}$ *p_{jx} : item difficulty index* (Baykul, 2015, p.229; Ray, Hundleby & Goldstein, 1962)	$+\infty \geq \alpha_{4j} \geq -2$ *Distribution of item scores; if the coefficient is 0, it is symmetrical; as it gets closer to +/-2, it moves away from symmetry (Baykul, 2015, p.229; Ray, Hundleby & Goldstein, 1962).*
Item Difficulty Index	It is the rate or percentage of correct answers to an item. It indicates ease-difficulty (Cohen & Swerdlik, 2009, p.159). It can also be seen as the probability of answering the item correctly (Baykul, 2015, p.219; Gulliksen, 1950, p.366).	Item Difficult Index (p_i) $p_i=\frac{N_D}{N}$ *N_D : maddeyi doğru yanıtlayan sayısı* *N : tüm yanıtlayıcıların sayısı* (Crocker & Algina, 2008, p.311; Gulliksen, 1950, p.366)	$0,20 \geq p_i \geq 0,00$ very difficult $0,40 \geq p_i \geq 0,21$ difficult $0,60 \geq p_i \geq 0,41$ moderate $0,80 \geq p_i \geq 0,61$ easy $1,00 \geq p_i \geq 0,81$ very easy *As the item difficulty index increases, the item becomes easier, and as it decreases, it becomes more difficult (Gulliksen, 1950, p.366).*

Psychometric Property	Definition	Estimation Method	Criterion
Item Discrimination Index	Item discrimination index to distinguish those who have the desired feature to be measured with the item and those who do not; since this feature of the item expresses the purpose of the measurement, the index obtained is also called the item validity coefficient (Crocker & Algina, 2008, p.313; Gulliksen, 1950, p.369).	Item Discrimination Index (r_{jx}) $r_{j(X\text{-}j)}=\frac{r_{jx}.S_x\text{-}s_j}{\sqrt{s_j^2+S_x^2\text{-}2.r_{jx}.S_x.s_j}}$ *S_x : standard deviation of test* (Baykul, 2015, p.242; Gulliksen, 1950, p.369)	$1{,}00 \geq r_{jx} \geq 0{,}40$ very good item $0{,}39 \geq r_{jx} \geq 0{,}30$ reasonably good but possibly subject to improvement $0{,}29 \geq r_{jx} \geq 0{,}20$ marginal items, usually needing and being subject to improvement $0{,}19 \geq r_{jx} \geq -1{,}00$ poor items, to be rejected or improved by revision *As the item discrimination index increases, the power of the item to reveal the difference between having and not having a feature increases (Ebel & Frisbie, 1991, p.232; ; Gulliksen, 1950, p.369).*
Item Distraction Index	It is a measure of how balanced the item distractors function together.	Item Distraction Index (ζ) $\zeta =1\text{-}\frac{\sqrt{\sum_{i=0}^{n}(f_i\text{-}\delta)^2}}{\sqrt{(q.n\text{-}\delta)^2+(a\text{-}1)(0\text{-}\delta)^2}}$	$0{,}00 \geq \zeta \geq 1{,}00$ *It can be said that as the ζ value increases, the distractors work together in a balanced way, and as the ζ value decreases, the distractors operate unevenly, that is, one or more distractors are inadequately designed.*
Item Reliability Index	It is the measure of the contribution of the item to the total score variance (Crocker & Algina, 2008, p.320; Gulliksen, 1950, p.375).	Item Reliability Index (r_j) $r_j = s_j.r_{jx}$ *r_{jx} : item score-test score correlation in which the item is included* (Crocker & Algina, 2008, p.320; Gulliksen, 1950, p.375)	$-0{,}50 \geq r_j \geq 0{,}50$ *As the r_j value increases, the contribution of the item to the test reliability also increases and decreases as it decreases (Crocker & Algina, 2008, p.320; Gulliksen, 1950, p.365).*
Item Validity Index	It is the measure of the contribution of the item to the criterion (fit/prediction) validity proof (Crocker & Algina, 2008, p.320).	Item Validity Index (ρ_j) $\rho_j = s_j.r_{jx}$ *r_{jy} : correlation between item score and external criterion score* (Crocker & Algina, 2008, p.320)	$-0{,}50 \geq \rho_j \geq 0{,}50$ *As the ρ_j value increases, the contribution of the item to the test validity also increases, and as it decreases, it decreases (Crocker & Algina, 2008, p.320).*

Psychometric Property	Definition	Estimation Method	Criterion
Content Validity Ratio	The content validity ratio is a measure based on expert opinions in providing evidence for the content validity of the items (Lawshe, 1975).	Content Validity Ratio (CVR) $\text{CVR}=\frac{N_N}{\frac{N}{2}}-1$ *N_N : Number of experts saying "necessary"* *N : total numbers of experts* (Lawshe, 1975)	$1{,}00 \geq \text{CVR}_j \geq -1{,}00$ *It was thought that as the value of the CVR increased, it represented the content better, and as it decreased, it could not.*
Item Factor Loading	It is the correlation of a variable with a factor (Khine, 2013, p.73).	Item Factor Loading (β) $Y_i=\beta_{i0}+\beta_{i1}F_1+\beta_{i2}F_2+\ldots+e_i$ *Y_i : variable i* *F_1 : first factor* *β_{i1} : Factor loading of Y_i variable on F_1 factor* *e_i : error* (Widaman, 1993)	*Factor loading values can be examined for statistical significance as a correlation value (Guilford, & Lyons, 1942, p.245). Since the load value of the items on the relevant factor was desired to be high, it was thought that as the β value increased, the item measured the structure better, and as it decreased, this level decreased.*
Differential Item Functioning	Bias is the systematic error of test scores of individuals in different subgroups depending on the group they belong to (Zumbo, 1999). To be able to say that an item is biased, it must first contain a Differential Item Function (DIF).	Mantel-Haenszel Statistics (Δ_{MH}) $\alpha_{MH}=\frac{\sum_{t=1}^{(L-1)}\frac{A_tD_t}{T_t}}{\sum_{t=1}^{(L-1)}\frac{B_tC_t}{T_t}}$ $\beta_{MH} = \ln(\alpha_{MH})$ $\Delta_{MH} = -2{,}35\,(\beta_{MH})$ (Mantel ve Haenszel, 1959)	$1{,}0 > \lvert\Delta_{MH}\rvert$ not DIF $1{,}5 > \lvert\Delta_{MH}\rvert \geq 1{,}0$ moderate DIF $\lvert\Delta_{MH}\rvert \geq 1{,}5$ high DIF (Zieky, 1993) *It is calculated based on the scores of male and female students.*

Appendix.2

Test Form	DIF Identified Items	Mean of Focal Group	Mean of Reference Group	MH	ΔMH	p
1	13	0,813	0,202	17,2085	-2,9040	0,01
1	19	0,730	0,326	5,5977	-1,7578	0,03
2	4	0,240	0,811	0,0736	2,6624	0,04
2	11	0,065	0,361	0,1231	2,1380	0,04
2	20	0,700	0,065	33,3123	-3,5781	0,00
3	2	0,485	0,876	0,1337	2,0538	0,04
3	7	0,909	0,576	7,3291	-2,0329	0,03
3	18	0,654	0,899	0,2115	1,5855	0,04
3	19	0,221	0,836	0,0556	2,9482	0,04
4	6	0,666	0,205	7,7059	-2,0840	0,03
5	1	0,655	0,152	10,5660	-2,4062	0,02
5	14	0,235	0,639	0,1735	1,7875	0,04
5	16	0,927	0,577	9,3073	-2,2767	0,02
6	3	0,272	0,886	0,0479	3,1010	0,04
6	9	0,576	0,072	17,4291	-2,9170	0,01
6	12	0,111	0,476	0,1378	2,0228	0,04
6	19	0,570	0,036	35,6501	-3,6473	0,00
7	8	0,791	0,079	44,0193	-3,8626	0,00
7	9	0,265	0,062	5,4316	-1,7271	0,03
7	13	0,809	0,270	11,4678	-2,4898	0,02
8	7	0,414	0,130	4,7233	-1,5845	0,03
8	8	0,086	0,316	0,2025	1,6298	0,01
8	16	0,633	0,101	15,2885	-2,7833	0,03
8	18	0,880	0,520	6,7471	-1,9484	0,04
9	4	0,675	0,219	7,4171	-2,0451	0,02
9	13	0,780	0,270	9,5719	-2,3053	0,03
9	16	0,935	0,302	33,1838	-3,5742	0,03
9	20	0,540	0,166	5,8945	-1,8105	0,02
10	8	0,254	0,040	8,0716	-2,1314	0,00
10	9	0,667	0,166	10,0491	-2,3550	0,03
10	12	0,269	0,816	0,0828	2,5422	0,02
11	1	0,408	0,131	4,5859	-1,5543	0,02
11	7	0,872	0,543	5,7547	-1,7861	0,04
11	8	0,086	0,887	0,0120	4,5166	0,03
12	15	0,849	0,316	12,1446	-2,5483	0,03
12	16	0,353	0,111	4,3787	-1,5072	0,04
13	5	0,107	0,408	0,1733	1,7888	0,02
13	17	0,650	0,177	8,6309	-2,1997	0,03
14	13	0,215	0,578	0,1997	1,6443	0,04
15	6	0,515	0,189	4,5657	-1,5498	0,02
15	9	0,661	0,115	15,0617	-2,7680	0,04
15	12	0,572	0,028	46,4808	-3,9181	0,03
16	3	0,315	0,058	7,4860	-2,0545	0,02
16	6	0,157	0,592	0,1284	2,0951	0,00
16	19	0,763	0,370	5,4983	-1,7395	0,03

Focal Group : Boys (n=2137); Reference Group : Girls (n=2005)

The ethics committee approval for this study was obtained from the Ethics Committee of the Rectorate of Hacettepe University, dated 10/09/2020 and numbered 35853172-300.

Statement of Contribution of Researchers to the Article:

1st author contribution rate: %50

2nd author contribution rate: %50

Conflict of Interest Statement:

There is no conflict of interest

Statement of Financial Support or Acknowledgment:

No financial support was received from any institution for this study. No Acknowledgment.