

The Effect of Incorrect Questions in Multiple Choice Mathematics Exams on the Psychometric Properties of the Test

Kevser Ağah (Teacher)

Ministry of National Education - Türkiye

ORCID: 0000-0003-0017-496X

kevseragah44@gmail.com

Assist. Prof. Dr. Azmi Türkan

Siirt University - Türkiye

ORCID: 0000-0003-2546-5122

azmiturkan@gmail.com

Abstract

The aim of this study is to determine and evaluate the effect of incorrect questions in multiple choice mathematics exams on the psychometric properties of the test in line with the opinions of the participants. In this study, convergent parallel mixed design was preferred since qualitative research data and quantitative data were collected together. In the quantitative dimension of research, an achievement test was applied to 100 eighth grade secondary school students attending three different public schools in the Southeastern Anatolia Region. The qualitative study part was conducted with 10 students with different success levels. Stratified sampling was used to collect quantitative data, and criterion sampling method was used to collect qualitative data. In the quantitative part of the study, the multiple choice mathematics achievement test prepared by the researchers was used. The achievement test consisted of a total of 30 questions, 10 at the easy level, 10 at the medium level, and 10 at the hard level, which were obtained in line with the opinions of experts working in the relevant field. After the achievement test application was completed, a semi-structured interview form developed by researchers was applied to participants. This form was used to determine the feelings and thoughts of the participants regarding the questions applied. Test Analysis Program (TAP) was used for the analysis of quantitative data. Content analysis was performed by creating codes, categories and themes for qualitative data analysis. When the findings were examined, it was seen that the presence of erroneous questions in multiple-choice tests positively affects item discrimination. Another finding obtained in the study is that the difficulty levels regarding the incorrect items were low. However, the participants stated that the incorrect questions among the questions made them think more and look at all the questions from different aspects with a questioning attitude.

Keywords: Multiple choice test, Mathematics, High-level cognitive features, Incorrect question



**E-International Journal
of Educational
Research**

Vol: 14, No: 1, pp. 139-153

Research Article

Received: 2022-10-01

Accepted: 2023-01-21

Suggested Citation

Ağah, K. & Türkan, A. (2023). The effect of incorrect questions in multiple choice mathematics exams on the psychometric properties of the test, *E-International Journal of Educational Research*, 14(1), 139-153. DOI: <https://doi.org/10.19160/e-ijer.1182923>

Extended Abstract

Problem: Evaluation in the context of mathematics lesson is the process of collecting evidence about students' mathematical knowledge, knowledge and skills, developing their mathematical knowledge and skills, understanding their mathematical goals in depth, providing learning-teaching, and determining students' tendency and interest in mathematics (NCTM, 2000). Looking at these functions, it can be assumed that among the purposes of assessment in mathematics these should be: To enable students to use mathematics in daily life and in complex situations and to determine how accurately they use it, to develop students in this context, to direct them to research and question, to determine and develop the ability of students to give the most logical answers by considering all possible solutions to mathematical problems (Pandey & Smith, 1991). Various measurement tools such as tests, inventories, rubrics or scales occupy an important place and are used in the assessment-evaluation stages (Tan, 2013). Tests from measurement and evaluation tools; written exam, oral exam, short answer test, true false test, multiple choice test and matched tests (Uyan, 2015). The multiple choice test is one of the most used test types in mathematics. One of the important limitations of multiple-choice tests is that they cannot measure high-level cognitive features (Üstüner & Şengül, 2014). This research aims to eliminate this limitation and thus enable the student to look at the question with a critical eye, expand his/her higher level thinking skills or increase his/her higher level cognitive steps. In the study, it was aimed to help individuals use their critical thinking and questioning skills actively by including the wrong answers in the mathematics multiple choice test exams and to distinguish between those who know and those who do not. When the relevant literature is examined, no study has been found in the context of Turkey regarding the effect of having erroneous questions in multiple-choice tests on the psychometric properties of the test. The aim of the study is to determine the effect of erroneous questions in multiple choice mathematics exams on the psychometric properties of the test and to evaluate them in line with the students' opinions. In line with this general purpose, answers were sought to the following questions.

1. Does the wrong question have an effect on the difficulty index of the test and the items?
2. Does the wrong question have an effect on the discrimination of the test and the items?
3. What are the students' views on the wrong questions?

Method: The research was designed according to mixed-pattern research models in which both qualitative and quantitative research methods are used together. Convergent design was preferred from mixed pattern research. Convergent pattern is expressed as studies in which qualitative and quantitative data are collected together and the data support each other (Creswell & Clark, 2014). Within the scope of this research, since qualitative research data and quantitative data were collected together, convergent parallel mixed design was preferred.

In the quantitative dimension of the research, an achievement test was applied to 100 eighth grade secondary school students attending three different public schools in the Southeastern Anatolia Region. The qualitative study part was carried out with 10 students with different success levels. Stratified sampling method was used to collect quantitative data. With the stratified sampling method, the population in the universe is divided into homogeneous groups and healthy data are obtained according to the determined conditions (Kılıç, 2013). For the collection of qualitative data, criterion sampling method was used. The criterion sampling method enables the determination of the sample by taking into account the predetermined criteria (Marshall & Rossman, 2014; Yıldırım & Şimşek, 2016).

In the quantitative part of the study, the multiple choice mathematics achievement test prepared by the researchers was used. The achievement test consisted of a total of 30 questions, 10 questions at easy level, 10 questions at medium level, and 10 questions at hard level, in line with the opinions received from experts in the relevant field. There are a total of 6 erroneous questions determined among 30 multiple-choice questions. The erroneous question is in the form of two questions from each level and has been prepared in such a way that the answer is not in the options.

In the qualitative part of the study, an observation form was created in which the reactions and behaviors of the students observed by the researchers while the achievement test was being applied were written. After the achievement test application was completed, a semi-structured interview form was applied to learn the thoughts and feelings of the students determined by the researcher.

Descriptive statistics were used for quantitative data analysis. Descriptive statistics are the best way to summarize, organize and interpret large numbers of data (Sözbilir, 2010). Test Analysis Program (TAP) was used for the analysis of quantitative data (Brooks & Johanson, 2003). The item difficulty index and item discrimination index of the problem were determined through TAP. Excel was also used to interpret these tests. Item discrimination power index is the feature of a test item that distinguishes students who are competent and inadequate in terms of knowledge (Kutlu, et al., 2008). Finding the item discrimination index in certain intervals can be interpreted differently. Content analysis was performed by creating codes, categories and themes for qualitative data analysis. The main purpose in content analysis is to collect similar verbal or written data around a category and theme, to examine and interpret it objectively and present it to the reader (Yıldırım & Şimşek, 2006). The obtained data were analyzed and divided into coherent and meaningful parts by using coding in a general framework.

Findings: When the findings are examined, it is seen that 24 items are not wrong questions and 6 items are deliberately prepared incorrectly. It has been observed that the 3rd, 5th and 28th items, which are not faulty items, should be removed from the test. When the difficulty values of the items that are not faulty, except for the items that should be discarded, are examined, it is observed that the items vary between 0.15 and 0.70. It was determined that item 13 of these items was very difficult, while the other items were found to be of medium difficulty. In addition, when the wrongly prepared items are examined, it is seen that the discrimination value of the 22nd item is below 0.19 and "very weak, it should definitely be removed", the 8th item is "quite good but still needs to be improved" and the other 4 items are "very good". item" is understood. When the item difficulty value for the wrong items is examined, it is understood that the 12th and 15th items are of medium difficulty but closer to the difficult level. It was determined that the other items from the wrong questions were at a very difficult level. In other words, it is understood that the non-erroneous items generally have medium difficulty and discrimination, while the faulty items are generally high in discrimination and the difficulty levels are very difficult. According to the findings, it was determined that the participants' test average score was 11.47 (out of 30 items), the average difficulty of the test was 0.38, the average discrimination was 0.50, the point biserial value was 0.48 and the Kr-20 value was 0.89. In other words, it is generally understood that the test is valid and reliable, with average difficulty and good discrimination level.

When the findings obtained in general terms are examined, it is seen that the presence of incorrect questions in multiple-choice tests affects item discrimination positively. Another finding obtained within the scope of the research is that the difficulty levels of the incorrect items were low. When the participant's views are examined, the fact that there are erroneous questions among the questions shows that they are pushed to think more, to look at all questions with a questioning attitude and from different aspects.

Suggestions: According to the research findings, it was concluded that consciously putting the wrong questions in the achievement tests decreased the item difficulty index, that is, the item also made the test more difficult, and it also increased the item discrimination. In this context, it is thought that conscious use of incorrect questions in competitive examination systems or in qualifying examinations in the recruitment process may be beneficial. Because, in competitive exams, individuals who know and do not know can be eliminated in a healthy way with items with high distinctiveness and low level of difficulty, that is, difficult. In this context, it is recommended that students be confronted with incorrect questions while choosing items with high difficulty levels in competitive exams that require elimination between participants.

When the participant's views are examined, the fact that there are erroneous questions among the questions shows that they are pushed to think more, to look at all questions with a questioning attitude and from different aspects. When 21st century skills are examined, one of the most important features expected from educational institutions is to raise individuals who think critically and question (Cansoy, 2018). This shows that having erroneous questions highlights students' critical thinking and questioning skills. Another conclusion we reached with the opinions of the participants is that the wrong question sometimes takes the student's time and reduces their motivation. In these cases, teachers can be offered to encourage and motivate the student.

Çoktan Seçmeli Matematik Sınavlarındaki Hatalı Soruların Testin Psikometrik Özelliklerine Etkisi ¹

Öğretmen Kevser Ağah
Milli Eğitim Bakanlığı - Türkiye
ORCID: 0000-0003-0017-496X
kevseragah44@gmail.com

Dr. Öğr. Üyesi Azmi Türkan
Siirt Üniversitesi - Türkiye
ORCID: 0000-0003-2546-5122
azmiturkan@gmail.com

Özet

Bu araştırmanın amacı çoktan seçmeli matematik sınavlarındaki hatalı soruların o testin psikometrik özelliklerine etkisini belirleyip katılımcıların görüşleri doğrultusunda değerlendirmektir. Bu araştırma kapsamında nitel araştırma verileri ile nicel veriler birlikte toplandığından yakınsayan paralel karma desen tercih edilmiştir. Yapılan araştırmanın nicel boyutunda Güneydoğu Anadolu Bölgesinde üç farklı devlet okuluna devam eden 100 sekizinci sınıf öğrencisine başarı testi uygulanmıştır. Araştırmanın nitel kısmı ise akademik başarı düzeyi farklı 10 öğrenciyle yürütülmüştür. Nicel verilerin toplanmasında tabakalı örnekleme, nitel verilerin toplanması için ise ölçüt örnekleme yöntemi kullanılmıştır. Yapılan araştırmanın nicel kısmında, araştırmacılar tarafından hazırlanan çoktan seçmeli matematik başarı testi kullanılmıştır. Başarı testi ilgili alanda çalışan uzman görüşleri doğrultusunda elde edilen kolay düzeyde 10, orta düzeyde 10 ve zor düzeyde 10 soru olmak üzere toplam 30 sorudan oluşmuştur. Araştırmanın nitel kısmında, başarı testi uygulandığı sırada araştırmacılar tarafından gözlemlenen katılımcıların tepki ve davranışlarının yazıldığı gözlem formu kullanılmıştır. Başarı testi uygulaması bittikten sonra araştırmacılar tarafından geliştirilen yarı yapılandırılmış görüşme formu katılımcıların uygulanan sorulara yönelik duygu ve düşüncelerini belirlemek için uygulanmıştır. Nicel verilerin analizi için Test Analiz Programı (TAP) kullanılmıştır. Nitel veri analizi için kod, kategori ve temalar oluşturularak içerik analizi yapılmıştır. Elde edilen bulgular incelendiğinde çoktan seçmeli testlerde hatalı soruların bulunmasının madde ayırt ediciliğini olumlu yönde etkilediği görülmüştür. Araştırmada elde edilen bir diğer bulgu ise hatalı olan maddelere ilişkin güçlük düzeylerinin düşük çıkmasıdır. Bununla birlikte katılımcılar, sorular arasında hatalı soruların olmasının kendilerini daha fazla düşünmeye, bütün sorulara sorgulayıcı tavırla ve farklı yönleriyle bakmaya ittiğini belirtmişlerdir.

Anahtar Kelimeler: Çoktan seçmeli test, Matematik, Üst düzey bilişsel özellikler, Hatalı soru



**E-Ululararasi
Eğitim Araştırmaları
Dergisi**

Cilt: 14, No: 1, ss. 139-153

Araştırma Makalesi

Gönderim: 2022-10-01
Kabul: 2023-01-21

Önerilen Atıf

Ağah, K. & Türkan, A. (2023). Çoktan seçmeli matematik sınavlarındaki hatalı soruların testin psikometrik özelliklerine etkisi, *E-Ululararasi Eğitim Araştırmaları Dergisi*, 14(1), 139-153. DOI: <https://doi.org/10.19160/e-ijer.1182923>

¹ Bu çalışma TÜBİTAK 2209-A Üniversite Öğrencileri Araştırma Projeleri Destekleme Programı tarafından desteklenmiştir, proje no: 1919B012000694

GİRİŞ

Eğitim öğretim süreçlerinin olmazsa olmaz parçalarından biri ölçme ve değerlendirmedir (Heritage, 2007). Bunun önemli nedenlerinin başında ölçme ve değerlendirmenin, eğitim öğretim süreçlerinde öğretmen, öğrenen, ortam, kurum ve programlar tarafından verilen kararların isabetliliğinden ve sürdürülebilirliğinden emin olmanın en önemli yollarından biri olması gelmektedir (Dokumacı-Sütçü ve Bulut, 2016a). Eğitimde ölçme; öğrenciyi bilişsel, duyuşsal ve psikomotor alanlarında inceleyerek, bu inceleme sonucunda elde edilen sonuçların sayı, sembol veya terimlerle belirtilmesidir (Turgut, 1986). Ölçme işleminde değişkenlerdeki farklılaşma önemli bir unsurdur (Doğan, 2019; Güler, 2011). Değerlendirme, ölçme sonucunun belirlenen kriterlere göre karşılaştırılarak bir sonuca varılması, yorumlanması, karar verilmesi (Lord ve Novick, 1968; Güler, 2011; Toprakçı, 2017). öğrenme seviyesini belirleme ve öğrenmenin gerçekleşip gerçekleşmediğini saptama süreci (Beevers ve Paterson, 2003) olarak ifade edilmektedir. Genel olarak öğretmenler; öğrenme-öğretme süreçlerini geliştirmek ve düzenlemek, gelecekteki planlarını gerçekleştirmek, öğrencilerinin öğrenme ihtiyaçlarını bilişsel, duyuşsal ve psikomotor davranışları adı altında analiz ederek onların öğrenmedeki değişme ve gelişmelerini izlemek, kazandırmış oldukları bilgi ve becerilerin ne düzeyde olduğunu belirlemek için ölçme-değerlendirmeden faydalanırlar (Smith, 2006; Semerci, 2011).

Matematik dersi bağlamında değerlendirme, öğrencilerin matematik bilgi, birikim ve becerileri ile ilgili kanıt toplama, matematiksel bilgilerini ve becerilerini geliştirme, matematiksel amaçlarını derinlemesine anlama, öğrenmeyi-öğretmeyi sağlama, öğrencilerin matematiğe olan eğilim ve ilgilerini belirleme sürecidir (National Council of Teachers of Mathematics [NCTM], 2000). Bu işlevlere bakılarak matematikteki değerlendirmenin amaçları; öğrencilerin günlük hayatta ve karmaşık durumlarda matematiği kullanmalarını sağlamak ve ne kadar doğru kullandıklarını belirleyerek bu bağlamda öğrenciyi geliştirmek, araştırma yapmaya ve sorgulamaya yönlendirmek, öğrencilerin matematik problemlerinde tüm olası çözümleri düşünerek en mantıklı cevapları verebilme becerilerini belirleyip geliştirebilmek şeklinde sıralanabilir (Pandey ve Smith, 1991).

Ölçme-değerlendirme aşamalarında testler, envanterler, puanlama anahtarları veya ölçekler gibi çeşitli ölçme araçları önemli bir yer tutmakta ve kullanılmaktadır (Tan, 2013). Matematik öğretiminde, her ne kadar çağdaş ölçme araçlarına yönelik olumlu tutum ve yönelim gösterilse de (Dokumacı-Simitci ve Bulut, 2016b) genellikle bilgi ve beceri düzeyindeki davranışların çoğunlukla geleneksel değerlendirme yaklaşımıyla ölçülmesi tercih edilmektedir (Birgin ve Baki, 2012; Çelikkaya vd. 2010). Geleneksel değerlendirme yaklaşımındaki ölçme yöntemlerinde öğrenci başarısının değerlendirilmesi, genellikle öğretim-öğrenim süreci hesaba katılmadan ve daha çok ürün ve sonuç odaklı bir şekilde ele alınmaktadır (Toptaş, 2011; Toprakçı, 2008). Geleneksel yaklaşımdaki test araçları: yazılı yoklama, sözlü yoklama, kısa cevaplı test, doğru yanlış testi, çoktan seçmeli test ve eşleştirmeli testlerdir (Uyan, 2015). Bunların içerisinde çoktan seçmeli test matematikte en çok kullanılan test türlerinden biridir.

Çoktan seçmeli testlerde, sorular öğrenciye olası cevaplarıyla (seçeneklerle) birlikte yazılı olarak verilir ve doğru seçeneğin işaretlenmesi istenir. Çoktan seçmeli testlerde kullanılan sorulara test maddesi denir. Test maddeleri genellikle iki temel kısımdan oluşur. Bunlar madde kökü (soru kökü) ve seçeneklerden; seçenekler de kendi içerisinde; doğru cevap (anahtar) ve çeldiricilerden oluşmaktadır (Doğruöz, 2022). Çoktan seçmeli testlerin genel özelliklerine bakıldığında en çok tercih edilen, çok fazla soruya yer verilebilme özelliğinden güvenilirlik ve geçerliliği yüksek olan test olduğu görülür (Temizkan ve Sallabaş, 2011). Güvenirliğin yüksek olma nedeni, madde sayısı artırılarak testin duyarlılığının artırılmasıdır. Geçerliliğin yüksek olmasının nedeni ise öğrencinin cevabını değerlendirirken yazı güzelliği, kompozisyon becerisi gibi öğretmeni etkileyen durumlarının olmamasıdır. Madde ayırtedicilik ve güçlük değerleri ile maddelerin testin amacına ne derece hizmet ettiği belirlenmektedir (Allen, 2012; DeVellis, 2006). Ayrıca çoktan seçmeli testlerin puanlaması objektif, kısa süreli ve uygulanması da kolaydır (Gelbal, 2013; Simonson vd., 2012).

Çoktan seçmeli testlerin güçlü yönlerinin yanı sıra çeşitli sınırlılıkları da vardır. Bu sınırlılıklar; öğrencinin yorumlamalarına kapalı olması, şans faktörü ve üst düzey bilişsel özellikleri ölçememe gibi özellikler sayılabilir (Baştürk, 2014; Race vd. 2005; Simkin ve Kuechler, 2005). Çoktan seçmeli testlerde cevaplayıcıya çeşitli seçenekler sunulduğu için cevaplayıcının kendi yorumunu katması engellenmektedir. Bu durum çoktan seçmeli testlerin sınırlılıklarından bir tanesidir (Başol, 2018; Tağrikulu vd., 2021). Şans

faktörü, bilenle bilmeyen arasındaki farkın doğru bir şekilde saptanmasını engelleyen bir unsurdur. Tek doğru cevabın olduğu test sorularında, bazen cevap hakkında bir fikri olmayan öğrenciler şanslarını deneyip soruyu cevaplarlar ve öğretmen cevabı doğru olarak değerlendirse de kimi zaman öğrencinin gerçekten bilip bilmediğinden emin olamaz. Bu durumda şans başarısını azaltmak için düzeltme formülü kullanılabilir ya da seçenek sayısını arttırılabilir ancak bu durum şans başarısını tamamen ortadan kaldırmaz (Yaman, 2016).

Çoktan seçmeli testlerin bir diğer sınırlılığı ise üst düzey bilişsel özellikleri ölçememesidir (Üstüner ve Şengül, 2014). Günümüzde eğitim öğretim faaliyetlerinin en önemli amaçları arasında öğrencilerin üst düzey düşünme ve araştırma-inceleme becerilerine katkı sağlamak yer almaktadır (Seheryeli, 2018). Benzer şekilde Boud (2007), bu tür sınavların üst düzey düşünme becerilerinden ziyade hatırlama seviyesindeki becerileri ölçmede kullanıldığını belirtmiştir. Üst düzey öğrenme becerilerine sahip bireyler, karşılaştıkları durumları içselleştirebilmekte ve edindiklerini farklı durumlarda kullanarak, olaylara yaratıcı ve eleştirel bakabilmektedirler (Filiz, 2013). Eğitim öğretim faaliyetlerinde gerek ulusal düzeyde geniş ölçekli sınavlarda gerek yerelde öğretmen yapımı sınavların birçoğunda çoktan seçmeli madde sınavların tercih edilmesi (Doğruöz, 2022) ve bu sınavların üst düzey düşünme becerilerini destekleyememesi büyük eksiklik olarak görülmektedir. Bireylerin üst düzey düşünme becerilerini harekte geçirebilmesi için onlara alışlagelmiş durumların dışına çıkma fırsatları sağlanabilir. Çoktan seçmeli maddelerden oluşan bir testte alışılmışın dışına çıkarak cevap şıkları arasında hatalı sorular bulundurmak, öğrenenlerin düşünme becerilerine katkı sağlayabileceği düşünülmektedir. Böylelikle çoktan seçmeli testlerin belirtilen üst düzey düşünme becerileri ölçebilme kısıtlılığının giderilmesine katkı sağlayabilecektir. Testler aracılığıyla bireylerin sahip oldukları beceri farklılıklarını göz önünde bulundurarak ölçmek için uygun zorluk düzeyinde ölçümlerin yapıldığı maddelerin seçilmesi önemlidir (Estrada, 2018). Bu çalışmada hatalı cevaplı soruları matematik çoktan seçmeli test sınavlarına dâhil ederek bireylerin eleştirel düşünme ve sorgulama becerilerini aktif şekilde kullanmaya teşvik etmek, bilen ile bilmeyen bireylerin ayırt edilmesini sağlamak hedeflenmiştir. Konu ile ilgili alanyazın incelendiğinde çoktan seçmeli testlerde hatalı soruların bulunmasının testin psikometrik özelliklerine etkisi ile ilgili herhangi bir çalışmaya rastlanmamıştır.

Araştırmanın amacı, çoktan seçmeli matematik sınavlarındaki hatalı soruların testin psikometrik özelliklerine etkisini belirleyip katılımcıların görüşleri doğrultusunda değerlendirmektir. Bu genel amaç doğrultusunda aşağıdaki sorulara da yanıtlar aranmıştır.

1. Hatalı sorunun testin ve maddelerin güçlük indeksine bir etkisi var mıdır?
2. Hatalı sorunun testin ve maddelerin ayırt ediciliğine bir etkisi var mıdır?
3. Hatalı sorulara ilişkin katılımcı görüşleri nelerdir?

YÖNTEM

1. Araştırma Deseni

Bu araştırma nitel ve nicel araştırma yöntemlerin bir arada kullanıldığı karma desen araştırma modellerine göre tasarlanmıştır. Karma desen araştırmalarından yakınsayan desen kullanılmıştır. Yakınsayan desen nitel ve nicel verilerin birlikte toplandığı ve verilerin birbirlerini destekler nitelikte olduğu araştırmalardır (Creswell ve Clark, 2014). Nicel araştırma, merak edilen olguları ve olayları gerekli koşulların sağlandığı ortamlarda; deneme, gözlem veya deney yoluyla nesnelleştirilerek sayılarla ifade edilmesidir (Öztürk, 2015; Tunç, 2020). Nitel araştırma, merak edilen bir olayın, algının, kişinin doğal ortamında kendini gerçekçi ve bütüncül bir şekilde ortaya konulmasına yönelik; gözlem, görüşme ve doküman analizi gibi nitel veri toplama yöntemleri kullanılarak, nitel bir sürecin izlendiği araştırmadır (Yıldırım ve Şimşek, 2016). Bu araştırma kapsamında da nitel araştırma verileri ile nicel veriler birlikte toplandığından yakınsayan paralel karma desen tercih edilmiştir.

2. Çalışma Grubu

Araştırmanın nicel boyutunda Güneydoğu Anadolu Bölgesinde üç farklı devlet okuluna devam eden 100 ortaokul sekizinci sınıf öğrencisine başarı testi uygulanmıştır. Çok sayıda maddenin bulunduğu ölçme araçlarına ilişkin geçerlik ve güvenirlik çalışması açısından testlerin psikometrik özelliklerinin

ölçüldüğü araştırmalarda örneklem sayısının en az 100 olması gerektiği belirtilmiştir (Kline, 2005; MacCallum vd., 1999). Nitel çalışma kısmı ise başarı düzeyi farklı 10 katılımcıyla yürütülmüştür. Nicel verilerin toplanmasında tabakalı örnekleme yöntemi kullanılmıştır. Tabakalı örnekleme yöntemi ile evrende bulunanlar homojen gruplara ayrılarak belirlenen koşullara göre sağlıklı verilerin elde edilmesi sağlanır (Kılıç, 2013). Bu bağlamda okullar Liselere Giriş Sınavı başarılarına göre tabakalara bölünmüştür. Bu tabakalar arasından, nicel verilerin toplanmasında Liseye Giriş Sınavı (LGS) sonuçlarına göre üç farklı başarı seviyesine sahip okul tercih edilmiştir. Nitel verilerin toplanması için ise ölçüt örnekleme yöntemi kullanılmıştır. Ölçüt örnekleme yöntemi, önceden saptanan ölçütlerin dikkate alınarak örneklemin belirlenmesini sağlar (Marshall ve Rossman, 2014; Yıldırım ve Şimşek, 2016). Nitel çalışma grubunun belirlenmesinde ölçütler farklı okulların olması ve bu okullarda uygulanan başarı testinde farklı başarı gösteren bireylerin olmasıdır.

Tablo 1’de katılımcıların grup dağılımlarının cinsiyete göre dağılımlarına yer verilmiştir. Katılımcılar 1., 2. ve 3. grup olmak üzere 3 farklı grup altında ele alınmıştır.

Tablo 1. Gruplar- Cinsiyet Dağılımı

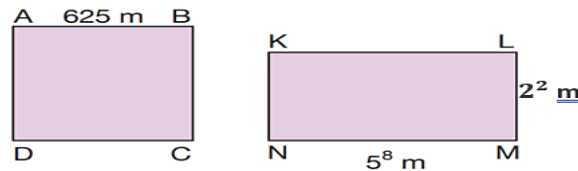
Grup	Cinsiyet		Toplam
	Kız	Erkek	
1. Grup	6	11	17
2. Grup	28	20	48
3. Grup	20	15	35
Toplam	54	46	100

Tablo 1 incelendiğinde Katılımcıların 54’ü kız iken 46’sı erkek öğrencilerden oluşmaktadır. Katılımcıların 17’si düşük, 48’i orta ve 35’i yüksek seviyede başarıya sahip okullardan seçilmiştir. Katılımcı seçiminde belirlenen okullardan veriler gönüllülük esasına göre toplanmıştır. Bazı başarı seviyesinden katılımcılar uygulamaya katılmakta gönüllü olmamalarından dolayı araştırmaya dahil edilmemiştir.

3. Veri Toplama Araçları

Araştırmanın nicel verilerinin toplanmasında, araştırmacılar tarafından hazırlanan çoktan seçmeli matematik başarı testi kullanılmıştır. Araştırma kasapmında testin psikometrik özellikleri belirlenerek teste ilişkin geçerlik ve güvenirlik değerleri belirlenmeye çalışılmıştır. Ancak testin geliştirilmesi aşamasında geçerlik ve güvenirliği saptamak için uzman görüşlerine başvurulmuştur. Nitekim Karasar (2010) ve Karp (2007) test hazırlama ve ölçek geliştirme çalışmalarında uzman görüşünün geçerlik ve güvenirliği arttıran bir unsur olduğunu belirtmişlerdir. Başarı testi, bir alan uzmanı ve bir ölçme ve değerlendirme olmak üzere iki uzman görüşü doğrultusunda çözümü kısa ve kolay olan 10 soru, çözümü orta düzeyde 10 soru, çözümü zor ve uzun zaman alan 10 soru olmak üzere toplam 30 sorudan oluşmuştur. 30 çoktan seçmeli soru arasında belirlenen toplam altı hatalı soru bulunmaktadır. Hatalı soru her düzeyden ikişer soru şeklinde olup ve cevabı seçeneklerde olmayacak şekilde hazırlanmıştır. Buradaki amaç katılımcıların farklı düzeydeki sorulara verdikleri tepkiyi belirlemektir. Şekil 1’de hatalı maddelere ilişkin örnek bir madde bulunmaktadır.

7. Şekildeki KLMN dikdörtgeninin alanının ABCD karesinin çevresine oranı kaç m’dir?



- A) 125 m
- B) 5^5 m
- C) 5^6 m
- D) 2^5 m

Şekil 1. Matematik Başarı Testi Hatalı Madde Örneği

Araştırmanın nitel kısmında, başarı testi uygulandığı sırada araştırmacılar tarafından gözlemlenen katılımcıların tepki ve davranışlarının yazıldığı gözlem formu kullanılmıştır. Gözlem, bir olayı, olguyu veya nesneyi belirli araştırma hedeflerine yönelik niteliklerini belirlemek için; doğal ortamında dikkatli ve planlı bir şekilde incelenmesidir (Çetin vd., 2010). Başarı testi uygulaması bittikten sonra araştırmacılar tarafından belirlenen katılımcıların düşünce ve duygularını öğrenmek için, yarı yapılandırılmış görüşme formu uygulanmıştır. Nitel araştırma tekniği olan yarı yapılandırılmış görüşme formu, araştırmacının önceden hazırladığı görüşme formuna ek olarak, görüşme sırasında cevaplar doğrultusunda formda olmayan soruları eklemesidir. Sınav sonrası katılımcıların hatalı soruyu gördüklerinde ne düşündükleri, soruyu nasıl çözmeyi planladıkları gibi soruların cevabına yarı yapılandırılmış görüşme formuyla ulaşılmıştır. Bu görüşme formu araştırmacılar tarafından hazırlanmıştır. Başarı testi ve yarı yapılandırılmış görüşme formu hazırlanırken bir alan uzmanı bir de ölçme ve değerlendirme uzmanı olmak üzere iki uzmandan görüş alınmıştır.

4. Verilerin Analizi

Nicel veri analizlerinde betimsel istatistiklere yer verilmiştir. Betimsel istatistik çok sayıda veriyi özetleme, düzenleme ve yorumlama için kullanılan en iyi yoldur (Sözbilir, 2010). Nicel verilerin analizi için Test Analiz Programı (TAP) kullanılmıştır (Brooks ve Johanson, 2003). TAP aracılığıyla teste bulunan soruların madde güçlük indeksi ve madde ayırt edicilik indeksi tespit edilmiştir. Test ve madde analizlerinin yorumlanmasında Excel Programından da yararlanılmıştır.

Madde ayırt edicilik indeksi, bir test maddesinin bilgi bakımından yeterli ve yetersiz öğrencileri birbirinden ayırması özelliğine denir (Kutlu vd., 2008). Madde ayırt edicilik indeksleri almış oldukları değer aralığına göre farklı şekilde yorumlanabilmektedir. Tablo 2’de madde ayırt edicilik indekslerinin değerlendirilmesine ilişkin değerler ve bu değerlere bağlı yorumlar bulunmaktadır.

Tablo 2. Madde ayırt edicilik indeksine bağlı maddenin değerlendirilmesi

Madde Ayırt Edicilik İndeksi	Maddenin Değerlendirilmesi
0,40 ve daha büyük	Çok iyi madde
0,30-0,39	Oldukça iyi ama yine de geliştirilebilir
0,20-0,29	Düzeltilmesi ve geliştirilmesi gerekir
0,19 ve daha küçük	Çok zayıf mutlaka çıkarılmalı

Kaynak: (Hasançebi vd., 2020)

Madde güçlük indeksi, test maddesinin öğrenciye göre zorluk derecesinin sayısal olarak hesaplanmasıdır. Güçlük indeksinin değeri 0 ile 1 arasındadır. Güçlük indeksi 1 değerine yaklaştıkça madde kolaylaşmakta, 0 değerine yaklaştıkça madde zorlaşmaktadır. Madde güçlük indeksinin 0,50 değerine yaklaşması da maddenin orta güçlükte olduğunu göstermektedir (Tekin, 2000). Madde güçlük indeksi değerinin 0,20’den küçük olması “çok zor”, 0,20 ile 0,40 arasında olması “zor” (0,20 ile 0,40 dahil), 0,41 ile 0,60 arası olması “orta”, 0,61 ile 0,80 arası olması “kolay” ve 0,80’den büyük olması “çok kolay” olduğu anlamına gelmektedir (Adıgüzel ve Özdoğru, 2013).

Nitel veri analizi için kod, kategori ve temalar oluşturularak içerik analizi yapılmıştır. İçerik analizinde temel amaç birbirine benzeyen sözel veya yazılı verileri belirli kategori ve temalar etrafında toplayarak verilerin objektif bir şekilde incelenmesi ve yorumlanması, okuyucuya sunulmasıdır (Yıldırım ve Şimşek, 2006). Elde edilen verilerin incelenerek, tutarlı ve anlamlı bölümlere ayrılması, genel bir çerçeve içerisinde kodlama kullanılarak yapılmıştır.

5. Araştırma Süreci

Verilerin toplanması için bir Devlet Üniversitesi Rektörlüğünden Etik Kurul izni alınmıştır. İlgili iznin alınmasından sonra verilerin toplanması için İl Milli Eğitim Müdürlüğünden uygulama izni alınmıştır. Veri toplama araçlarının geliştirilmesi için öncelikle alan ve ölçme ve değerlendirme alan uzmanı aracılığıyla başarı testi oluşturulmuştur. Başarı testine uygun gözlem formu ve yarı yapılandırılmış görüşme formları oluşturulmuştur. Verilerin toplanması üç aşamadan oluşmuştur. İlk aşamada öğrenciye çoktan seçmeli matematik testi uygulanıp yanlış sorularda öğrencinin tepkisi gözlemlenmiştir. Sonraki aşamada ise test sonuçları analiz edilerek öğrencilerin verdiği cevaplar doğrultusunda sayısal verilere ulaşılmıştır. Son

aşamada ise öğrencilerle görüşülüp fikir, düşünce, duygu ve tepkileri hakkında yarı yapılandırılmış görüşme formu ile araştırmacı kontrolünde veriler toplanmıştır.

Çoktan seçmeli madde uygulamalarında katılımcılara, test maddelerini içten ve gönüllülük esasına göre doldurulması için gerekli açıklamalar yapılmıştır. Ayrıca maddeler arasında hatalı madde olduğu bilgisi verilmemiştir. Katılımcıların uygulama esnasında diğer katılımcılardan etkilenmemeleri için sınav esnasında maddeler hakkında herhangi bir sesli yorum yapmamaları varsa yorumları bu yorumlarını sınav kâğıdına yansıtmaları gerektiği ile ilgili bilgilendirme yapılmıştır.

BULGULAR

1. Nicel Verilere İlişkin Bulgular

Tablo 3'te katılımcıların 30 maddelik çoktan seçmeli test maddelerine ilişkin madde analiz sonuçlarına yer verilmiştir. Madde verileri, madde güçlük indeksleri ve madde ayırt edicilik indeksleri doğrultusunda analiz edilmiştir.

Tablo 3. Hatalı soru bulunan test maddelerine ilişkin madde analiz sonuçları

Soru	Madde Güçlüğü (P_j)	Madde Ayırt ediciliği (r_{jx})	Soru	Madde Güçlüğü (P_j)	Madde Ayırt ediciliği (r_{jx})
1	0,54	0,61	16	0,67	0,59
2	0,59	0,65	17	0,60	0,62
3	0,94	0,16	18	0,26	0,62
4	0,51	0,65	19	0,39	0,57
5	0,04	0,04	20	0,45	0,77
6	0,59	0,81	21*	0,15	0,52
7*	0,19	0,49	22*	0,05	0,17
8*	0,16	0,35	23	0,42	0,70
9	0,30	0,36	24	0,31	0,43
10	0,45	0,74	25	0,44	0,33
11	0,53	0,51	26	0,23	0,32
12*	0,34	0,70	27	0,20	0,41
13	0,15	0,25	28	0,10	0,18
14	0,70	0,33	29	0,54	0,81
15*	0,28	0,69	30	0,35	0,66

* sembolü olan maddeler hatalı hazırlanmış maddeleri göstermektedir.

Tablo 3'te katılımcıların 30 maddelik çoktan seçmeli test maddelerine ilişkin madde analiz sonuçlarına yer verilmiştir. Madde verileri, madde güçlük indeksleri ve madde ayırt edicilik indeksleri doğrultusunda analiz edilmiştir.

Tablo 3 incelendiğinde 24 maddenin hatalı soru olmadığı, 6 maddenin bilinçli olarak hatalı bir şekilde hazırlanmış maddeler olduğu görülmektedir. Hatalı olmayan maddelerden 3., 5. ve 28. maddelerin testten atılması gerektiği gözlemlenmiştir. 13.maddenin ise düzeltilmesi ve geliştirilmesi gereken bir madde olduğu belirlenmiştir. Atılması gereken maddeler dışında hatalı olmayan maddelere ilişkin güçlük değerleri incelendiğinde ise maddelerin 0,15 ile 0,70 arasında değiştiği gözlemlenmiştir. Bu maddelerden 13. maddenin çok zor diğer maddelerin orta güçlükte olduğu tespit edilmiştir.

Tablo 3'te hatalı bir şekilde hazırlanan maddeler incelendiğinde, bu maddelerden 22. maddenin ayırt edicilik değerinin 0,19'un altında olduğu için "çok zayıf mutlaka çıkarılmalı", 8. maddenin "oldukça iyi ama yine de geliştirilmesi" gerektiği diğer 4 maddenin ise "çok iyi madde" olduğu anlaşılmaktadır. Hatalı maddelere ilişkin madde güçlük değeri incelendiğinde 12. ve 15. maddelerin orta güçlükte ancak zor düzeye daha yakın oldukları anlaşılmaktadır. Hatalı sorulardan diğer maddelerin çok zor düzeyde oldukları tespit edilmiştir. Başka bir ifade ile genel anlamda hatalı olmayan maddelerin genellikle orta güçlükte ve ayırt ediciliğinin olduğu, hatalı maddelerin ise genellikle ayırt etme gücünün yüksek olduğu güçlük düzeylerinin ise çok zor seviyede olduğu anlaşılmaktadır.

Tablo 4'te, teste ilişkin analiz bulgularına yer verilmiştir. Tabloda kişi sayısı, test ortalama puanı, test ortalama güçlük değeri, test ortalama ayırt ediciliği, nokta çiftserili değeri ve Kr-20 değerine yer verilmiştir.

Tablo 4. Hatalı soru bulanan teste ilişkin test analiz bulguları

Grup	N	Ortalama	Test Ortalama Güçlüğü	Test Ortalama Ayırt ediciliği	Nokta Çiftserili	Kr-20
Hatalı Soru Bulunan Başarı Testi	100	11,47	0,8	0,50	0,48	0,89

Tablo 4 incelendiğinde katılımcıların toplam 30 puan üzerinden hesaplanan test puanlarının, test ortalama puanının 11,47 (30 madde üzerinden), testin ortalama güçlüğü 0,38 olduğu, ortalama ayırt ediciliğin 0,50 olduğu, nokta çiftserili değerinin 0,48 ve Kr-20 değerinin 0,89 olduğu tespit edilmiştir. Başka bir ifade ile genel olarak araştırmanın amacına bağlı olarak, testin ortalama güçlükte ve iyi ayırt edicilik seviyesine sahip olduğu anlaşılmaktadır.

2. Nitel Verilere İlişkin Bulgular

Araştırmanın nitel verilerine ilişkin bulgular kısmında genel olarak yarı yapılandırılmış görüşmeye ve gözlem raporlarına ilişkin verilere yer verilmiştir. Yarı yapılandırılmış görüşme analizleri sonucunda katılımcıların "Soruları Tekrar Gözden Geçirme ve Zaman" ve "Çözümün Sorgulanmasına İlişkin Eylem" temaları altında toplandığı tespit edilmiştir. "Soruları Tekrar Gözden Geçirme ve Zaman" teması altında "Hatalı soruları tekrar çözme" ve "Hatalı sorularda zaman harcama" kodları toplanmıştır.

Araştırmada kullanılan yarı yapılandırılmış görüşme formunun analizleri sonucunda bazı katılımcılar, çoktan seçmeli testlerde hatalı soruların bulunmasını kendileri açısından teste daha kuşku ile yaklaştıklarını, hatalı soruları tekrar çözme denediklerini ve farklı çözüm yolları kullandıklarını belirtmişlerdir.

"Hatalı soruyu çözerken farklı yollar da kullandım. iki defa soruyu tekrardan çözmeyi denedim ama şıklarda cevaba ulaşamadım." (Ö1)

"18. soru bana göre hatalıydı ya da ben çözemedim. İlk başta şıklardan gittim farklı yollar denedim, olmadı." (Ö9)

Katılımcıların büyük çoğunluğu hatalı sorulardan dolayı fazla zaman harcadıklarını ve soruları çözerken zorlandıklarını ayrıca telaşlandıklarını dile getirmişlerdir. Katılımcılardan soru çözümünde zorlanan katılımcılara ait bazı görüşler:

"7. soru da zorladım, cevabı bulamadım. Hatalı olup olmadığından da emin olamadım. Yine de çözümüne en yakın cevabı buldum. Ondan sonra 21. soruda da zorlandım. Çok işlem yapmama rağmen cevabı bulamadım. Ayrıca 22.soruda da zorlandım..... Çok yanlış çıkardığımı düşündüm ve telaşlandım. Ayrıca zamanımı da aldı." (Ö2)

"7,22 ve 23. Sorularda zorlandım. Hatalı olduklarını düşündüğüm için zorlandım." (Ö8)

"7. soruda zorlandım. Bir de 22 ve 23. işlemleri fazla geldi ve sorularda zor geldi." (Ö9) şeklindedir.

Katılımcıların hatalı sorunun fazla zaman alma ile ilgili görüşlerinden bazıları:

"Zamanımı aldı. Doğru şıkkı bulmaya çalıştım. Biraz da motivasyonumu düşürdü." (Ö3)

"Sadece zaman kaybım oldu." (Ö8)

"Zamanımı aldı ve motivasyonumu düşürdü." (Ö5) şeklindedir. Burada Ö3 ile Ö5 katılımcıları motivasyonlarının düştüklerini de belirtmektedirler.

"Çözümün Sorgulanmasına İlişkin Eylem" teması altında ise bireylerin "Kendi bilgisine güven" ve "Yanlış çözdüğünü düşünme" kodlarının bulunduğu tespit edilmiştir.

Bu durumla ilgili "Hatalı soruyu görünce yanlış çözdüğünü mü ya da sorunun hatalı olduğunu mu düşündün?" sorusuna ilişkin analiz, elde edilen farklı bulgulardan dolayı ikiye ayrılmıştır. Kendi bilgisine, çözüm yoluna güvenip sorunun hatalı olduğunu düşünen bazı katılımcıların görüşleri:

"Bildiğim çözüm yoluyla cevap çıkmayınca çok şaşırdım. Nasıl böyle olabilir diye düşündüm. Daha sonra sorunun hatalı olduğunu düşündüm." (Ö3)

"Hatalı soruyu görünce öğretmene bildirdim. Şaşırdım aynı zamanda. İlk başta kendi yanlış çözdüğümü düşündüm. Farklı bir yoldan da cevap çıkmayınca sorunun hatalı olduğunu düşündüm." (Ö4)

"Kesinlikle sorunun hatalı olduğunu düşündüm." (Ö7) şeklindedir.

Bilgisine ve çözüm yoluna güvenmeyip soruyu yanlış çözdüğünü düşünen katılımcılardan bazılarının görüşleri:

"Benim hatalı çözdüğümü düşündüm. Soruda hata olabileceğini düşünmedim." (Ö5)

"Benim yanlış çözdüğümü düşündüm." (Ö9) şeklindedir.

Araştırmada kullanılan gözlem formu raporlarının analizi sonucunda, katılımcıların çoktan seçmeli başarı testi uygulamasının başlamadan önce ve başladığı sırada tepki ve davranışlarında farklı bir durum gözlemlenmemiştir. Uygulamanın birkaç dakika sonrasında birkaç katılımcının ilk hatalı soruyu fark edip el kaldırdıklarını ve araştırmacılara bildirdiklerini, birkaç katılımcının "bu soru hatalı" diye belirttiğini, birkaç katılımcının ise çözmeye devam ettiğini gözlemlenmiştir. Katılımcılardan bazıları diğer hatalı sorularla karşılaştıklarında ise araştırmacıya tekrar belirtmek yerine sorunun yanına not aldıkları gözlemlenmiştir. Sınav bittikten sonra bazı katılımcılar araştırmacıya birkaç hatalı sorunun olduğunu söylemiş ve cevaplarının ne olduğunu sormuşlardır.

Araştırmacılar tarafından incelenen başarı testlerinde, hatalı soruları fark eden bazı katılımcıların hatalı sorunun yanına işaret koyduğu, hatalı diyerek kendi çözümünü belirttiği, testin boş kısımlarında soru ile ilgili yorumlarını yazdıkları, bunun yanı sıra birkaç katılımcının da hatalı olmayan soruya hatalı soru şeklinde yaklaştıkları tespit edilmiştir. Ayrıca katılımcıların soruyu fark edemeyen katılımcıların ise genelde boş bıraktıkları veya çözümlerine yakın seçeneği işaretledikleri gözlemlenmiştir.

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu araştırmada, çoktan seçmeli matematik testlerinde kasıtlı olarak bulundurulmuş hatalı maddelerin testin psikometrik özelliklerine etkisi ve katılımcıların hatalı maddelere yönelik görüşleri incelenmiştir. Eğitim sistemlerinde uygulanacak testler hazırlanırken planlı ve aşamalı bir yol izlenmelidir. Baykul'a (2015) göre bir test geliştirilirken, teste ilişkin amaç, kapsadığı konu, teste ilişkin maddelerin belirlenmesi, maddelerin gözden geçirilmesi, pilot uygulama, asıl uygulamanın puanlanması yapılması, maddelere ilişkin analizleri yapılması, maddelerin seçilmesi ve testin son haline ilişkin istatistiklerin yapılması şeklinde sıralanmaktadır. Testin amacı bu süreçte büyük ölçüde önemlidir. Diğer aşamalar testin amacı etrafında şekilleneceği için bu amaç itina ile tespit edilmelidir (Baykul, 2015). Elde edilen bulgular incelendiğinde çoktan seçmeli testlerde hatalı maddelerin bulunmasının madde ayırt ediciliğini olumlu yönde etkilediği görülmüştür. Başka bir ifade ile testte bulunan hatalı maddelerin bilen ve bilmeyen katılımcıları birbirinden ayırt edebildiği tespit edilmiştir. Bir ölçme aracı bulunduran maddelerden beklenen en önemli özelliklerden bir tanesi bilen ile bilmeyenlerin birbirinden ayırt edilmesidir (Baykul, 2015; Crocker ve Algina, 2006; Hasançebi vd., 2019; Saraç, 2018). Benzer şekilde Allen (2012) ve DeVellis (2006) madde ayırt edicilik ve madde güçlük işlemlerinin madde güvenilirlik ve geçerliği için önemli unsurlar olduğunu dile getirmişlerdir. Özellikle çoktan seçmeli testlerde şans faktörü varken bireyler arasındaki elde edilmiş kazanım farkının nitelikli bir şekilde ortaya çıkması önemli bir durumdur. Bu araştırmada elde edilen bu bulgunun bireylerin birbirinden ayırt edilmesinde kullanılabilecek bir yapı sunması bakımından önemli olduğu görülmektedir. Oluşturulacak bu yapının ulusal ve uluslararası sınavlarda, merkezi ve yerel ölçme değerlendirmelerde, özel ve resmî kurum ve kuruluşların işe alım sınavlarında kullanılması nitelik farkının sağlıklı bir şekilde ortaya çıkarılabilmesi için avantaj sağlayabileceği söylenebilir. Bireyler hakkında bir kanaate varılan bu tarz büyük ölçekli sınavlarda isabetli kararların alınması bireylerin gelecekleri ile ilgili

önemli bir belirleyicidir (Mor-Dirlik, 2021). Nitekim, Baykul (2010) ve Erkuş (2003); isabetli kararların alınmasında madde ayırt edicilik ve güçlük gibi istatistiksel işlerin yapılmasının gerekliliğine vurgu yapmışlardır. Bu bağlamda elde edilen bulgular hatalı maddelerin isabetli kararlar alma noktasında başvuru temel özelliklerden biri olan madde ayırt edicilik indeksini olumlu yönde etkilediğini göstermektedir.

Araştırmada hatalı olan maddelerin güçlük düzeylerinin düşük çıkması elde edilen bir diğer bulgudur. Madde güçlüğü düşük çıkan maddelerin zor maddeler olduğu katılımcıların büyük çoğunluğu tarafından doğru cevaplanmadığını göstermektedir (Tekin, 2000). Araştırma kapsamında elde edilen bu bulgu hatalı soruların katılımcıları zorladığı ve çoğunluğunun soruyu doğru cevaplamadığını göstermiştir. Günümüzde rekabete dayalı merkezi sınavlardaki uygulanan testlerin asıl amacı katılımcıları seçmek, sınıflamak ve yerleştirmektir. Bu tür sınavlarda bağıl değerlendirmeler kullanılmaktadır. Ulusal düzeyde yapılan sınavlarda bağıl değerlendirmeler rekabet unsurunu ön plana çıkarmaktadır (Duman, 2011). Bu sebeptendir ki Baykul (2015), madde güçlük indeksinin ve ayırt ediciliğin günümüzdeki ulusal ve uluslararası sınav sistemlerinde öğrenciyi seçmeye ve yerleştirmeye yönelik yani daha çok rekabete ve elemeye dayalı sınavlarda önemli bir unsur olduğunu belirtmektedir.

Bireylerin genellikle herhangi bir kurum veya kuruluşlarda, işe alım süreçleri kapsamında rekabetin olması için eleme ve seçmeye dayalı yöntemlerle bireylere çeşitli testler uygulanır (Türkben, 2019). Benzer şekilde Cohen ve Swerdlik (2018) güç testlerinde bulunan maddelerin zorluk düzeylerinin yüksek olduğunu belirtmiştir. Uygulanan testlerde madde güçlük indeksi ne kadar düşük olursa yani maddeler dolayısıyla test ne kadar zor olursa eleme amacıyla yapılan ölçme süreci de bir o kadar amaca hizmet eder ve verimli seçimler olur. Araştırma bulgularına göre hatalı soruların başarı testlerine bilinçli olarak koyulmasının madde güçlük indeksini düşürdüğüne yani maddeyi aynı zamanda da testi zorlaştırmasının yanında madde ayırt ediciliğini de yükselttiği görülmüştür. Çünkü rekabete dayalı olan sınavlarda bilen ile bilmeyen bireylerin sağlıklı bir şekilde elenebilmesi ayırt ediciliği yüksek, güçlük düzeyi düşük yani zor maddeler ile sağlanabilecektir. Çetin (2019) ölçme araçlarının en ideal madde güçlüğüne sınavın amacından amacına göre değişebileceğini belirtmiştir. Araştırmacı ayrıca çok başvurunun olduğu ama az kişinin alınması gereken rekabete dayalı sınavlarda ideal madde güçlüğüne zor maddelerden oluşan bir sınav olması gerektiği dile getirmiştir. Bu bağlamda rekabete dayalı, katılımcılar arası eleme yapılması gereken sınavlarda güçlük düzeyi yüksek maddeler tercih edilirken öğrencilerin hatalı sorularla karşı karşıya bırakılması önerilebilir.

Katılımcı görüşleri incelendiğinde, test maddeleri arasında hatalı maddelerin olmasının öğrencileri daha fazla düşünmeye sevkettiği, test maddelerine sorgulayıcı bir bakış açısıyla ve farklı yönleriyle bakmayı sağlayabileceği düşünülmektedir. 21. yüzyıl becerileri incelendiğinde eğitim kurumlarından beklenen en önemli özelliklerin başında eleştirel düşünen ve sorgulayan bireyler yetiştirilmesi gelmektedir (Cansoy, 2018). Eğitim öğretim faaliyetlerinde sıklıkla kullanılan çoktan seçmeli testlerin, bu bireylerin üst düzey düşünme becerilerine katkı sağlayabilecek şekilde dizayn edilmesi gerekmektedir (Brookhart, 2010). Bu beklenti paralelinde testlerde hatalı maddelerin kullanılmasının öğrencilerin eleştirel düşünme ve sorgulama becerilerini ön plana çıkardığını göstermektedir. Geleneksel ölçme araçları yerine çağdaş ölçme araçlarının tercih edilmesinin temel gerekçesi öğrencilerin üst düzey düşünme ve bilişsel becerilerinin geliştirilmesidir (Atabek Yiğit ve Balkan Kıyıcı, 2018; Duban ve Küçükylmaz, 2008; Orhan, 2007; Tuncel ve Kazu, 2019). Elde edilen bulgular çoktan seçmeli testlerde hatalı maddelerin olmasının öğrencilerin üst düzey düşünme ve bilişsel becerilerine katkı sağladığını göstermektedir. Katılımcı görüşlerinden elde edilen başka bir sonuç ise hatalı sorunun bazı durumlarda öğrencinin zamanını alıp motivasyonunu düşürdüğü yönündedir. Geleneksel ölçme araçlarından biri olan çoktan seçmeli testlerin öğrencilerin mantıksal düşünme ve öğrenme süreçlerinin bütüncül olarak algılamasını sağlayamamakta (Ben-Hur, 2006) buna karşın birçok ölçme değerlendirme işlemlerinde objektifliği sağlaması açısından (Baker, 2001; Özdemir, 2022) bu tür sınavlar kullanılmaktadır. Elde edilen bulgulara göre çoktan seçmeli testlerde bulunan bu eksiklik bilinçli bir şekilde hatalı şıkların bulunduğu maddeler ile bertaraf edilmesine katkı sağlayabileceği düşünülmektedir.

Yerel ölçme – değerlendirme, Milli Eğitim Bakanlığına bağlı okullarda eğitim alan öğrencilere, öğretmenler tarafından uygulanan ölçme değerlendirme faaliyetleridir (Özdaş, 2019). Okulda yapılan çoktan seçmeli sınavlarda hatalı maddelerin teste dahil etmesinin öğretmen ve öğrenci açısından birçok

olumlu etkileri olacağına inanılmaktadır. Öğrenci açısından bu olumlu etkilerin çoktan seçmeli soruya sorgulayıcı ve eleştirel bakabilmenin yanında üst düzey bilişsel basamağında ilerleme kaydetme ya da geliştirme, kendi yanıtını bulduğunda bilgisine güvenme ve bunun getirdiği başarı hissi olduğu söylenebilir. Öğretmen açısından olumlu etkileri ise çoktan seçmeli sınavlar için şans faktörünü minimuma indirme ve bilenle bilmeyenin arasındaki farkın daha kolay ayırt edilebilmesi olarak düşünülebilir. Aynı zamanda hatalı maddelerin bulunduğu sınavlarda yerel düzeydeki ölçme değerlendirmede kullanılmasının faydalı olacağı düşünülmektedir.

Rekabate dayalı sınavlara öğrencilere gerekli açıklamalar yapılarak, madde ayırdediciliğini arttırmak için cevap şıkları arasına kısa cevap yapabilecekleri seçenekler koyulması önerilebilir. Böylelikle sorulan sorunun sadece seçeneklerden ibaret olmadığı, öğrenciyi eleştirel bir bakış açısıyla düşünmeye sevk edileceği düşünülmektedir. Başka çalışmalarda, ilgili başarı testinin farklı şartlar altında uygulanarak standardizasyonun sağlanması çalışmaları yapılabilir. Öğrenci görüşlerine göre düşünme becerilerine katkı sağladığı görüşünden yola çıkarak farklı çalışmalarda öğrencinin hatalı soruların bulunduğu test başarıları ile üst düzey düşünme becerileri arasındaki ilişkinin tespit edilmesi önerilebilir.

KAYNAKÇA/REFERENCES

- Adıgüzel, O. C., & Özdoğru, F. (2013). Üniversitelerde ortak zorunlu yabancı dil I dersine yönelik bir akademik başarı testinin geliştirilmesi. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, 3(2).
- Allen, D. D., (2012). Validity and Reliability of the Movement Ability Measure: A Self-Report Instrument Proposed for Assessing Movement Across Diagnoses and Ability Levels. *Physical Therapy*, 87(7), 899-916.
- Atabek Yiğit, E. & Balkan Kıyıcı, F. (2018). Bilişsel stiller ve alternatif ölçme ve değerlendirme araçlarının kullanım yeterliği arasındaki ilişkinin incelenmesi. *Scientific Educational Studies*, 2 (1) , 39-55.
- Baker, F. B. (2001). *The basics of item response theory* (2nd ed.). College Park, MD: University of Maryland, ERIC Clearinghouse on Assessment and Evaluation
- Başol, G. (2018). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Baştürk, S. (2014). Çoktan seçmeli testler. S. Baştürk (Ed.), *Eğitimde ölçme ve değerlendirme* (s. 119-154). Nobel Akademik Yayıncılık.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması* (2. Baskı). Pegem Akademi
- Baykul, Y. (2015). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Pegem Akademi.
- Beevers, C. E. & Paterson, J. S. (2003). *Automatic assesment of problem solving skills in mathematics active learning in higher education*. 4(2), 127-145. <https://doi.org/10.1177%2F1469787403004002002>
- Ben-Hur, M. (2006). *Concept-rich mathematics instruction. association for supervision and curriculum evelopment*, Alexandria, Virginia, USA.
- Birgin, O. & Baki, A. (2012). Sınıf öğretmenlerinin ölçme-değerlendirme uygulama amaçlarının yeni matematik öğretimi programı kapsamında incelenmesi. *Eğitim ve Bilim*, 37(165), 152-167
- Boud, D. (2007). Reframing assessment as if learning were important. In D. Boud, & N. Falchikov (Ed.), *Rethinking assessment in higher education: Learning for the longer term*, London: Taylor & Francis Group.
- Brookhart, S. M. (2010). *How to Assess Higher-Order Thinking Skills in Your Classroom*. Virginia: ASCD.
- Brooks, G. P., & Johanson, G. A. (2003). Test analysis program. *Applied psychological measurement*, 27, 305-306. <https://doi.org/10.1177%2F0146621603027004007>
- Cansoy, R. (2018). Uluslararası çerçevelere göre 21. yüzyıl becerileri ve eğitim sisteminde kazandırılması. *İnsan ve Toplum Bilimleri Araştırmaları Dergisi*, 7(4), 3112-3134. <https://doi.org/10.15869/itobiad.494286>
- Cohen, R.J. and Swerdlik, M.E. (2018). *Psychological testing and assessment: An introduction to tests and measurement* (9. Baskı). New York: McGraw-Hill Education.
- Crocker, L. M., & Algina, J. (2006). *Introduction to classical and modern test theory* (2nd edth) .Victoria: Thomson Wadsworth (ed.)
- Çelikkaya, T., Karakuş, U. & Öztürk Demirbaş, Ç. (2010). Sosyal bilgiler öğretmenlerinin ölçme-değerlendirme araçlarını kullanma düzeyleri ve karşılaştıkları sorunlar. *Ahi Evran Üniversitesi Eğitim Fakültesi Dergisi*, 11(1), 57-76.
- Creswell, J. W., & Plano-Clark, V. (2014). *Karma yöntem araştırmaları* (Çev Edt: Dede, Y. ve Demir, SB), Anı Yayıncılık, Ankara.

- Çetin, B. (2019). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı Yayıncılık.
- Çetin, T., Kuş, Z., & Karatekin, K. (2010). Sınıf ve sosyal bilgiler öğretmenlerinin gezi-gözlem yöntemine ilişkin görüşleri. *Sosyal Bilimler Araştırmaları Dergisi*, 158-180.
- DeVellis, R. F., (2006). Quantitative and Issues and Approaches: Classical Test Theory (CTT) and Item Response Theory (IRT). *Medical Care*, 44(11), 50-59
- Doğan, N. (2019). *Eğitimde ölçme ve değerlendirme. İnsan Davranışları ve Ölçme Araçlarını Sınıflama*. Ankara: Pegem Akademi Yayıncılık.
- Doğruöz, E. (2022). Çoktan seçmeli testler. *Uygulama örnekleri ile etkili üniversite eğitimi*. (Ed. Yöntem, M. K.) Eğitim Yayınevi.
- Dokumacı Sütçü, N., & Bulut, İ. (2016a). Ortaokul matematik öğretmenlerinin alternatif ölçme ve değerlendirme tekniklerine ilişkin yeterlik algıları ve bu teknikleri kullanma düzeyleri. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, 290. <https://doi.org/10.14582/DUZGEF.706>
- Dokumacı-Sütçü, N., & Bulut, İ. (2016b). Ortaokul Matematik Öğretmenlerinin Alternatif Ölçme ve Değerlendirme Tekniklerini Kullanma Düzeylerine İlişkin Yeterlik Algılarının Değerlendirilmesi/Assesment of Competency Perceptions of Secondary School Math Teachers about the Level of the Use of AMET. *e-Uluslararası Eğitim Araştırmaları Dergisi*, 6(3).
- Duban, N., & Küçükyılmaz, E. A. (2008). Sınıf öğretmeni adaylarının alternatif ölçme-değerlendirme yöntem ve tekniklerinin uygulama okullarında kullanımına ilişkin görüşleri. *İlköğretim online*, 7(3), 769-784
- Duman, B. (2011). Sınıf öğretmeni adaylarının bağıl değerlendirmeye ilişkin görüşleri. *NWSA-E Journal of New World Sciences Academy*, 6(1), 536-548
- Erkuş, A. (2003). Psikometri üzerine yazılar. Türk Psikologlar Derneği Yayınları.
- Estrada, E. (2018). Speeded test. In *The SAGE encyclopedia of educational research, measurement, and evaluation*, 1, s. 846-848). California
- Filiz, Y. S. (2013). *Öğrenme Öğretme Kuram ve Yaklaşımları*. Ankara: Pegem Akademi.
- Gelbal, S., (2013). *Ölçme ve Değerlendirme*. Selahattin Gelbal (Ed.). Eskişehir: Anadolu üniversitesi Web-Ofset.
- Güler, N. (2011). *Eğitimde Ölçme ve Değerlendirme*. Ankara: PegemA.
- Hasançebi, B., Terzi, Y., & Küçük, Z. (2020). Madde güçlük indeksi ve madde ayırt edicilik indeksine dayalı çeldirici analizi. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, 10(1), 224-240. <https://doi.org/10.17714/gumusfenbil.615465>
- Heritage, M. (2007). What do teachers need to know and do? *Phi Delta Kappan*, 89 (2), 140 – 145. <https://doi.org/10.1177%2F003172170708900210>
- Karasar, N. (2010). *Bilimsel araştırma yöntemi: kavramlar, ilkeler, teknikler*. Ankara: Nobel Yayınevi
- Karip, E. (Ed.). (2007). *Ölçme ve değerlendirme*. Pegem A Yayıncılık.
- Kılıç, S. (2013). Örneklem yöntemleri. *Journal of Mood Disorders*, 3(1), 44-6. <https://doi.org/10.5455/jmood.20130325011730>
- Kline, R. B. (2005). *Principles and practice of structural equation modeling (2nd ed.)*. New York, NY: Guilford
- Kutlu, Ö., Doğan, C.H., Karakaya, İ. (2008). *Öğrenci Başarısının Belirlenmesi: Performansa ve Portfolyoya Dayalı Durum Belirleme*. Ankara: PegemA Yayıncılık.
- Lord, F. & Novick, M. (1968). Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley.
- MacCallum, R. C., Widaman, K. F., Zhang, S., ve Hong, S. (1999). Sample Size in Factor Analysis. *Psychological Methods*, 4, 84-99
- Marshall, C. & Rossman, G. B. (2014). Designing qualitative research. Sage.
- Mor Dirlik, E. (2021). Farklı kuramlar kapsamında hesaplanan madde ayırt edicilik parametrelerinin karşılaştırılması. *Trakya Eğitim Dergisi*, 11(2), 732-744. <https://doi.org/10.24315/tred.700445>
- National Council for Teachers of Mathematics[NCTM]. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.
- Orhan, T. A. (2007). *Fen eğitiminde alternatif ölçme ve değerlendirme yöntemlerinin ilköğretim öğretmen adayı, öğretmen ve öğrenci boyutu dikkate alınarak incelenmesi*. Yayınlanmamış Doktora Tezi, Ankara: Gazi Üniversitesi, Eğitim Bilimleri Enstitüsü.
- Özdaş, F. (2019). Merkezi yerleştirme sınav sistemine ilişkin öğretmen ve öğrenci görüşlerinin değerlendirilmesi. *Mukaddime*, 10(2), 688-707 <https://doi.org/10.19059/mukaddime.509244>
- Özdemir, A. (2022). Ayırık Seçenekli Çoktan Seçmeli Testlerin Uygulanabilirliği. (Yayınlanmamış Doktora Tezi) Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

- Öztürk, H. (2015, Nisan 7). *Nitel ve nicel araştırma nedir? Aralarındaki farklar nelerdir?* <https://haldunozturk.com/nitel-ve-nicel-arastirma-nedir-aralarindaki-farklar-nelerdir/>
- Pandey, T. & Smith, T. R. (1991). A sampler of mathematics assessment (Eric Ed: 341 553). URL:www.eric.com
- Race, P., Brown, S., & Smith, B. (2005). 500 tips on assessment. (Second Ed.), Abingdon: Taylor & Francis Group.
- Saraç, H. (2018). Fen bilimleri dersi 'Maddenin değişimi' ünitesi ile ilgili başarı testi geliştirme: *Geçerlik ve güvenilirlik çalışması*. Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 18 (1), 416-445. <https://doi.org/10.17240/aibuefd.2018.-388815>
- Seheryeli, M. Y. (2018). *Yazılı Anlatım Becerisi Puanlama Anahtarının Güvenirliğinin Klasik Test, Genellenebilirlik ve Madde Tepki Kuramlarına Göre İncelenmesi* (Yüksek Lisans Tezi). Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Semerçi, Ç. (2011). Ölçme ve değerlendirme. E. Karip (Ed), Eğitimde ölçme ve değerlendirme. Pegem Akademi Yayıncılık.
- Simkin, M.G., & Kuechler, W.L. (2005). Multiple - choice tests and student understanding: What is the connection? *Decision Sciences Journal of Innovative Education*, 3 (1),73-97.
- Simonson, M., Smaldino, S., Albright, M. ve Zvacek, S. (2012). *Teaching and learning at a distance: Foundations of distance education* (3rd ed.). New Jersey: Prentice Hall.
- Smith S. S. (2006). *Early childhood mathematics* (3rd Ed.). Pearson Education
- Sözbilir, M. (2010). *Betimsel İstatistik*. <https://avys.omu.edu.tr/storage/app/public/yusuf.sisman/128524/6.Not.pdf>
- Tağrikulu, P., Bekdemir, Ü., & Omca Çobanoğlu, E. (2021). Üretim, dağıtım ve tüketim öğrenme alanına yönelik akademik başarı testi geliştirme: geçerlik ve güvenilirlik çalışması. *Kırşehir Eğitim Fakültesi Dergisi*, 22(3), 1984-2004. DOI: 10.29299/kefad.953899
- Tan, Ş. (2013). *Öğretimde ölçme ve değerlendirme KPSS el kitabı*. Pegem A Yayıncılık. 8. Baskı
- Tekin, H. (2000). *Eğitimde ölçme ve değerlendirme*. Yargı Yayınevi.
- Temizkan, M., & Sallabaş, M. E. (2011). Okuduğunu anlama becerisinin değerlendirilmesinde çoktan seçmeli testlerle açık uçlu yazılı yoklamaların karşılaştırılması. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, (30), 207-220.
- Toprakçı, E. (2018). *Sınıfa Dayalı Yönetim*. (1. Baskı). Ankara: Pegem Yayınları
- Toprakçı, E. (2017). *Sınıf Yönetimi*. (3. Baskı). Ankara: Pegem Yayınları
- Toptaş, V. (2011). Sınıf öğretmenlerinin matematik dersinde alternatif ölçme ve değerlendirme yöntemlerinin kullanımı ile ilgili algıları. *Eğitim ve Bilim*, 207.
- Tuncel, T., & Kazu, İ. Y. (2019). Ortaöğretim matematik öğretim programlarının ölçme ve değerlendirme boyutunda öğretmen görüşleri açısından incelenmesi. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 29(2), 163-179. <https://doi.org/10.18069/firatsbed.549200>
- Tunç, T. (2020). *Nicel Araştırma Yöntemi ve Aşamaları* <https://www.bilgiustam.com/nicel-arastirma-yontemi-ve-asamalari/>
- Turgut, M. F. (1986). *Eğitimde ölçme ve değerlendirme*. Saydam Yayınları.
- Türkben, E. (2019). *İşe alım sürecinde uygulanan seçme yerleştirme tekniklerinin iş tatmini ve işten ayrılma niyetine etkisi ve bir araştırma* (Doctoral dissertation, Marmara Üniversitesi (Türkiye)).
- Uyan, N. (2015, Aralık 8). *Geleneksel ölçme ve değerlendirme yöntemleri* <https://www.slideshare.net/NuranUyan/geleneksel-Imevedeerlendirmeyntemleri>
- Üstüner, A. & Şengül, M. (2004). Çoktan seçmeli test tekniğinin Türkçe öğretimine olumsuz etkileri. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 14(2), 197-208.
- Yaman, S. (2016). Çoktan seçmeli madde tipleri ve fen eğitiminde kullanılan örnekleri. *Gazi Eğitim Bilimleri Dergisi*, 2 (2) , 151-170
- Yıldırım, A. & Şimşek, H. (2006). *Sosyal bilimlerde nitel araştırma yöntemleri*, (5. Baskı), Seçkin Yayıncılık.
- Yıldırım, Y. & Şimşek, H. (2016). *Sosyal Bilimlerde Nitel Araştırma Yöntemleri*. Seçkin Yayıncılık.