Cumhuriyet International Journal of Education, 14(3): 688-700, 2025
DOI: https://dx.doi.org/10.30703/cije.1569140

Cumhuriyet International Journal of Education

| cije.cumhuriyet.edu.tr | Founded: 2011 Available online, ISSN: 2147-1606 Publisher: Sivas Cumhuriyet Üniversitesi

The Effect of Teaching Activities and Homework Assignments Based on Different Question Types on Student Achievement[#]

Mehmet Ali Kılıç[1,a,*], Ramazan Gürbüz[2,b]

[1]*Ministry of Education, Batman, Türkiye*
[2]*Faculty of Education, Adıyaman University, Adıyaman, Türkiye*
**Corresponding author*

Research Article

Acknowledgment
[#] This study is a part of the doctoral thesis.

History
Received: 18/10/2024
Accepted: 27/02/2025

This paper was checked for plagiarism using iThenticate during the preview process and before publication.

Copyright © 2017 *by Cumhuriyet University, Faculty of Education. All rights reserved.*

ABSTRACT

In this study, the effect of the question type used in mathematics teaching at the seventh grade level in secondary school and the quality of homework (Multiple Choice (MC), Open ended questions (OE) and Mixed Test (MT) questions and worksheets) on the mathematics achievement of students was examined. The research design used in this study is called "Alternative-Treatments Design with Pretest". The study group consisted of 78 students (G1=27, G2=28, G3=23). Different methods were used for each of the study group; in the first group, ES was used, in the second group, OE was used, and in the last group, a Mixed Test (MT) item-oriented teaching process was followed. In this context, parallel to the in-class teaching activities, the groups were assigned homework assignments. As data collection tools in the study, MC and OE tests developed by the researcher were applied as pre-test and post-test. The data were analyzed with SPSS 20 program. As a result of the study, it was observed that both the OE and the MC significantly increased the achievement of the students in all three groups. There was no significant difference between the post-test scores of the students in the MC data. In the OE test data, the group trained on the OE question type was significantly more successful than the group trained on the MT question type.

Keywords: Question types, homework, mathematics achievement

Farklı Soru Tiplerine Göre Gerçekleştirilen Öğretim Etkinliklerinin ve Ev Ödevlerinin Öğrenci Başarısına Etkisi[#]

Bilgi
[#]Bu çalışma doktora tezinin bir parçasıdır.
**Sorumlu yazar*

Süreç
Geliş: 18/10/2024
Kabul: 27/02/2025

Bu çalışma ön inceleme sürecinde ve yayımlanmadan önce iThenticate yazılımı ile taranmıştır.

Copyright

This work is licensed under Creative Commons Attribution 4.0 International License

ÖZ

Bu çalışmada, ortaokul yedinci sınıf düzeyinde matematik öğretiminde kullanılan soru tipi ve buna bağlı ev ödevlerinin niteliğinin (Çoktan Seçmeli (ÇS), Açık Uçlu (AU) ve Karma Test (KT) soruları ve çalışma yaprakları) AU, ÇS ve KT sorularının öğrencilerin matematik başarıları üzerindeki etkisi incelenmiştir. Bu çalışmada kullanılan araştırma deseni "Ön Testli Alternatif Müdahaleler Deseni" (Alternative-Treatments Design with Pretest) olarak belirlenmiştir. Araştırmanın çalışma grubu (G1=27, G2=28, G3=23) 78 öğrenciden oluşmaktadır. Çalışma gruplarının her birinde farklı yöntemler kullanılmıştır; ilkinde ÇS, ikincisinde AU son grupta ise Karma Test (KT) madde odaklı bir öğretim süreci takip edilmiştir. Bu kapsamda gruplar sınıf içi öğretim etkinliklerine paralel olarak hazırlanan ev ödevleri ile görevlendirilmişlerdir. Çalışmada veri toplama araçları olarak, araştırmacı tarafından geliştirilen Çoktan Seçmeli Test (ÇST) ve Açık Uçlu Test (AUT) ön test ve son test olarak uygulanmıştır. Verilerin analizi SPSS 20 programı ile yapılmıştır. Araştırma sonucunda her üç grupta da hem AUT hem de ÇST'nin öğrencilerin başarılarını anlamlı bir şekilde arttırdığı görülmüştür. ÇST verilerinde öğrencilerin son test puanları arasında anlamlı bir fark oluşmamıştır. AU test verilerinde ise AU soru tipi üzerinde eğitim yapılan grup KT soru tipi üzerinde eğitim yapılan gruptan anlamlı ölçüde daha başarılı bulunmuştur.

Anahtar Kelimeler: Soru tipleri, ev ödevleri, matematik başarısı

[a] *mehmetalikilic21@gmail.com* *https://orcid.org/0009-0005-5085-9950* | [b] *gurbuz@outlook.com* | *https://orcid.org/0000-0002-2412-5882*

How to Cite: Kılıç, M. A., & Gürbüz, R. (2025). Farklı soru tiplerine göre gerçekleştirilen öğretim etkinliklerinin ve ev ödevlerinin öğrenci başarısına etkisi. *Cumhuriyet International Journal of Education, 14*(3), 688-700.

Giriş

Eğitim programı hedef, içerik, eğitim durumları ve değerlendirme olmak üzere birbiriyle dinamik bir ilişki içerisinde olan dört temel öğeden oluşmaktadır (Arslan vd., 2018; Wolf vd., 2007). Eğitim sürecinin sonucunda programda belirlenen kazanımlara ne düzeyde ulaşıldığı değerlendirme çalışmalarıyla belirlenir (Baykul, 2010; Van den Akker, 2013). Bu anlamda değerlendirme öğesi eğitimin paydaşlarına dönüt sağlayarak eğitim sürecine dinamik (Turgut ve Baykal, 2014; Baykul, 2010; Öksüz ve Demir, 2019) ve döngüsel (National Council of Teachers of Mathematics [NCTM]), 1995) bir yapı kazandırır. Değerlendirme aynı zamanda, eğitimin niteliğini artıran en önemli bileşendir (Gültekin, 2011; National Resarcher Council [NRC], 2001). Eğitim bir bütün olarak değerlendirildiğinde bir alanında yapılan değişiklikler diğer alanlarda da gerçekleştirilmezse yapılan değişiklikler işlevsel olmayacaktır. Bundan dolayı ölçme değerlendirme süreçlerinin de diğer süreçlere paralel olarak değişmesi gerekir. Ancak özellikle merkezi sınavların eğitim sistemi üzerindeki baskısı değişimin tüm boyutlara yansımasına engel teşkil edebilmektedir(Birgin, 2010). Diğer bir ifadeyle, okullarda verilen eğitimin ulusal sınavlara göre şekillenmesi ölçme ve değerlendirme sürecinin Milli Eğitim Bakanlığı(MEB) reformlarına paralel gelişmesine mani olmaktadır (Birgin, 2010; Demirtaşlı, 2010).

Ölçme ve değerlendirme sürecinde kullanılan ölçme araçları amaca yönelik; çoktan seçmeli, açık uçlu ve yarı yapılandırılmış (MEB, 2015) olmak üzere farklılık göstermektedir. AU sınavlar ile ÇST arasında hangisinin tercih edileceği sınav uygulayıcılarının ihtiyaçları doğrultusunda karar verilmesi gerekmektedir. AU soru tiplerinin avantajları maddelerinin yazımı kolay olması, öğrencilerin üst düzey zihinsel becerilerini ölçebilecek yapıya sahip olması(Herman vd., 1994), kolay hazırlanması ve öğretmenin öğrencilerin kullandıkları çeşitli teknikleri değerlendirmesine olanak tanımasıdır (Chaoui, 2011). Dezavantajı ise puanlanmasının adil olamayacağı endişesi olarak değerlendirilmektedir. Benzer şekilde ÇS soru tiplerinin avantajları; puanlama kolaycılığı ve geniş bir kitlede uygulamak tek oturumda çok sayıda kişiyi test etme, yanıtlarının puanlanmasının kolay (Bible vd., 2008), objektif olması ve testin kapsam geçerliliğinin nispeten yüksek olması olarak sıralanabilir (Norman,1995; Demirtaşlı, 2010; Özçelik, 2011). Çoktan seçmeli sınavlar eğitimde en yaygın kullanılan soru tiplerindendir(Schuwirth & Van der Vleuten, 2004). ÇST dezavantajları ise, nitelikli soru yazımının zor olması (Demirtaşlı, 2010), ÇS testlerin hazırlanmasında güvenilirlik analizi için büyük bir örnekleme ihtiyaç duyulmasıdır. Bir diğer sınırlılık ise günlük hayat durumlarında dahi, çok boyutlu düşünemeyen sadece belli bir noktaya eğilen, 21. Yüzyıl becerileri olan farklı ve esnek düşünce gibi becerileri kazanamayan öğrencilerin yetişmesine neden olmaktadır(Birgili, 2014;Demirtaşlı, 2010; Gültekin ve Çıkrıkçı- Demirtaşlı, 2012; Özçelik, 2011, Satankus, 2016). Söz konusu sınırlılıkların ötesinde çoktan seçmeli soru tiplerinin yaygın kullanımı "çoktan seçmeli eğitime" sebep olmaktadır (Berberoğlu, 2010). Yapılandırmacı eğitim anlayışı ile örtüşmeyen ölçme yöntemi olduğu düşünülen ÇS soru tipleri, söz konusu anlayışın öğrenciye kazandırmak istediği becerileri ölçecek türden olmadığına inanılmaktadır (Anderson, 1998; Pohman, 1999; Berberoğlu, 2009).

Öğrencilerin bireysel farklılıklarından dolayı tek tip ölçme araçlarının tüm öğrencilere hitap edemeyeceği bundan dolayı farklı ölçme araçları ile öğrencilerin değerlendirilmesi gerektiği düşünülmektedir (Karaçam, 2009). NCTM (1995) sınıf içi değerlendirmelerde çoklu soru tiplerinin kullanılmasının öğrencilerin farklılıklarını yansıtabileceği düşünülmektedir. Öğrenciler, öğrenme kazanımlarının ölçüldüğü yöntemle öğrenme eğilimindedirler (Struyven vd., 2006; Akt., Koyuncu ve Özkan, 2019; Çepni, 2009; Beberoğlu vd., 2010; MEB,2011; Bilgeç, 2016). Milli Eğitim Bakanlığı ulusal sınavlarda sorulan ÇS soru tiplerinin öğrencilerin üst düzey düşünme becerileri arasında yer alan yaratıcılık, eleştirel düşünme ve problem çözme becerilerin ölçülmesinin (MEB,2021) zor olduğunu bundan dolayı; ulusal sınavlarda açık uçlu soruların sorulmasının alt yapısını oluşturmak amacıyla ABİDE projesini geliştirmişlerdir. ABİDE olarak kısaltılan "Akademik Becerilerin İzlenmesi ve Değerlendirilmesi" projesi, MEB tarafından iki yıllık peryotlar şeklinde; 4, 8 ve 10. Sınıfa devam eden öğrencilerin türkçe, matematik ve fen bilimleri dersinde edindikleri bilgi ve becerileri, ulusal düzeyde değerlendiren bir çalışmadır. MEB'in 32304 sayılı yönetmenliğine göre, "Okullarda uygulanan bütün sınavların yanıtların öğrenciler tarafından verildiği, farklı bilişsel düzeyde kazanımları ölçmeye yönelik madde içeren yazılı yoklama şeklinde yapılır" ifadesiyle ölçme değerlendirme sürecinde, açık uçlu soru tiplerinin tercih edilmesi zorunluluğu getirmiştir(MEB, 2023). Yeni uygulamanın sonuçlarını ise zaman gösterecektir.

Soru tipinin öğrenci başarısına etkisinin incelendiği araştırmalara bakıldığında, Eren (2015) yaptığı çalışma ile ÇS ve KT madde türlerini ölçen iki ölçme aracının puanlarını birbiriyle ilişkisinin yanında öğretmen ve öğrencilerin bu madde türleriyle ilgili görüşleri incelenmiş, hem öğretmen hem de öğrencilerin ÇS test formatını daha fazla tercih ettikleri gözlenmiştir. Benzer sonuçları Demir (2010) Türkiye'deki öğrencilerin matematik testlerinde ÇS soru tiplerini AU soru tiplerine tercih ettiklerini, aynı zamanda bu soru tiplerine dayalı ölçümlerde daha başarılı olduklarını bulgusuna ulaşmıştır. Özkan ve Özarslan (2018), yaptıkları çalışmada Programme for International Student Assessment(Uluslararası Öğrenci Değerlendirme Programı,)PISA 2003-2012 döneminde yapılan sınavın soru tiplerine göre öğrenci cevaplarının incelendiği çalışmada, öğrenciler ÇS soru tiplerinde daha yüksek başarı gösterirken, Karmaşık Çoktan Seçmeli ve AU soru tiplerinde ÇS soru tipine göre daha düşük başarı göstermişlerdir.

Toker ve Green (2016) Trends in International Mathematics and Science Study (Uluslararası Matematik

ve Fen Eğilimleri Araştırması) TIMSS 2011 verileri üzerinde yaptıkları çalışmada ÇS madde türlerinin Türk öğrenciler için daha kolay maddeler olduğunu belirlemişlerdir. Öğrencilerin genel puanlarını düşüren AU madde tiplerinin olduğu tespitine yer verilmiştir. PISA ve TIMSS gibi sınavlarda AU soru tiplerindeki başarının ÇS soru tipindeki başarıdan daha düşük olmasının sebebi Türk öğrencilerin bu sınavlarda yer verilen çoktan seçmeli madde dışında soru tiplerine aşina almamalarından kaynaklandığı (Demir, 2010), öğrencilerin deneyimleri ve aşina oldukları soru tiplerinde performanslarının daha yüksek olduğu gözlenmiştir (Ben-Simon ve Cohen 2004) .

Koğar ve Koğar (2019) başarısı yüksek öğrencilerin açık uçlu maddeler, başarısı düşük öğrencilerin çoktan seçmeli madde türlerinde daha başarılı olduğu görülmüştür. Aynı şekilde İlhan vd. (2020) TIMSS 2015 verileri üzerinden yürüttükleri çalışmada öğrencilerin açık uçlu maddelerde çoktan seçmeli maddelere kıyasla daha çok zorlandıklarını ve sekizinci sınıf düzeyindeki maddelerde güçlük indeksinin %25'inin maddenin türü (çoktan seçmeli–açık uçlu) ile açıklanabildiği belirlenmiştir. Bu oran dördüncü sınıf düzeyinde ise %7 olarak bulunmuştur. Bu çalışmalar, merkezi sınavın yapıldığı sınıf düzeyine yaklaşıldıkça öğrencilerin ÇS madde türlerindeki başarılarının artıp AU madde tipindeki başarılarının düştüğünü göstermektedir.

Cambaz (2000), çoktan seçmeli ve klasik türdeki sınavların matematik öğretimi üzerindeki etkisini karşılaştırmayı amaçlayan bir araştırmayı ilköğretim öğrencileri üzerinde yapmıştır. Çalışmada klasik sınavlarla değerlendirilen öğrencilerin, çoktan seçmeli sınavlarla değerlendirilen öğrencilerden matematik başarısı bakımından anlamlı bir şekilde yüksek çıkmıştır.

Türkiye'de yapılan sınavlarda öğrencilerin ÇS soru tipine maruz bırakılmaları ve söz konusu soru tiplerinde deneyim kazanması, ÇS soru tiplerinde başarılarını yükselten bir neden olabileceği düşünülmektedir (Demir, 2010;Üney, 2023). Öğrencilerin girecekleri ulusal nitelikteki sınavlarda karşılaşacakları soru türleri, çoktan seçmeli olmasından dolayı, okullarda bu soru tiplerine ağırlık verildiği görülmüştür (Demirtaşlı, 2012) benzer şekilde matematik öğretmenlerinin de ulusal düzeyde yapılan sınavlara uygun soruları tercih ettikleri görülmüştür (Yıldırım vd., 2012). Üney, (2023)'de yaptığı çalışmada 8. Sınıf öğrencilerinin çoktan seçmeli soru tipindeki başarısı, karma test ile hazırlanmış teste göre anlamlı ölçüde, çoktan seçmeli test lehine farklılık görülmüştür. Ülkemizde merkezi sınavlarda yalnızca çoktan seçmeli soru tipinin kullanılması, günümüz dünyasında hem uluslararası sınavlarda tercih edilen soru tipine hem de eğitim seviyelerinin yüksek olduğu ülkelerin ulusal alanda gerçekleştirdikleri geniş ölçekli madde türleri ile uyuşmamaktadır (Koyuncu ve Özkan, 2019). Öğrenciler eğitim süresince sınav odaklı çalışmaktalar (Çepni, 2009) değerlendirme sürecinde karşılaşılan soru tipleri, öğrencilerin öğrenme yöntemini şekillendirmektedir (Birgin, 2010; Arslan, 2016). Merkezi sınavlarda sorulan çoktan seçmeli sorular, ortaokul ve orta öğretim sistemini bu sınavların etkisi altına almıştır(Demirtaşlı,2010). Bundan dolayı okullarda yapılan eğitim faaliyetleri sınavda başarılı olma kaygısıyla, çoktan seçmeli soru çözme faaliyeti bir öğretim yöntemi olarak değerlendirilmektedir(Berberoğlu vd., 2010), bundan dolayı öğrenciler konuları kavramak yerine soru çözümünü sağlayacak algoritmaları ezberlemektedirler (Gültekin ve Demirtaşlı, 2012). Okuldaki ders ortamında ve öğrencilere verilen ev ödevlerinde, "soru çözme (çözdürme)" eğitim etkinliği olarak tercih edilmektedir(Konak, 2013).

İlgili literatür tarandığında, yapılan çalışmalarda matematik dersinde madde tipinin öğrenci başarısına etkisinin araştırıldığı, öğrenci başarılarının soru tiplerine göre değerlendirildiği çalışmaların bazılarında veri kaynağı TIMSS ve PISA gibi uluslararası merkezi sınavlar(Demir, 2010; Çelen, Çelik ve Seferoğlu, 2011; Aydın vd., 2012; İskenderoğlu ve Baki, 2011; Toker ve Green, 2016; Bozkurt, 2016; Özkan ve Özarslan, 2018; Koyuncu ve Özkan, 2019; İlhan vd.,) olduğu gözlenmişken bazıları ise ulusal çaptaki (Eren, 2015; Gültekin ve Demirtaşlı, 2012; Aydın ve Önder, 2010; Karaçam, 2009; Kastner ve Stangly, 2011) verilerden faydalanmıştır. Madde tiplerinin psikometrik özelliklerine göre karşılaştırıldığı araştırmalar da bulunmaktadır (Kastner and Stangl, 2010; Temizkan ve Sallabaş, 2011; Küçüker, 2017; Öksüz ve Demir, 2019). Deneysel araştırmalar ise sınırlı sayıda olup (Karaçam, 2009, Demir, 2011, Aydın ve Önder, 2010; Satankus, 2016), Aydın ve Önder, (2010) yaptıkları çalışmada, üniversite öğrencilerinin final sınavı hazırlığında kullandıkları ÇS ve AU soru tipine ilişkin deneysel bir çalışma yapmışlardır; ancak çalışma grubu üniversite öğrencileri olup eğitimin alt düzeylerine odaklanılmış bir çalışmaya rastlanmamıştır. Aynı zamanda gruplarda farklı sınav türü kullanılmamıştır. Bu çalışma eğitim kademesi olarak ortaokul düzeyinde olup AU, ÇS ve KT tipindeki soru yapılarının başarı belirleme düzeyleri üzerinde yapılmış bir çalışma olmasıyla alandaki boşluğu kısmen dolduracağı değerlendirilmektedir.

Çalışmanın Amacı

Bu araştırmanın genel amacı, ortaokul matematik dersinde soru türlerine göre öğrencilere verilen ev ödevlerinin ve gerçekleştirilen öğretim etkinliklerinin öğrencilerin KT, AUT ve ÇST'ye göre matematik başarılarına etkisini araştırmaktır. Bu amaçla aşağıdaki sorulara cevap aranmıştır:

1. Matematik dersinde ÇS, AU ve KT uygulanan grupların AU ve ÇS test puanlarına göre ön test ve son test puanları arasında fark var mıdır? Varsa bu fark anlamlı mıdır?

2. Matematik dersinde ÇS, AU ve KT uygulanan grupların AU ve ÇS test puanlarına göre başarı puanları arasında anlamlı fark var mıdır?

Yöntem

Araştırmanın Modeli

Bu çalışma kullanılan araştırma desenin literatürdeki adı "Ön Testli Alternatif Müdahaleler Deseni" (Alternative-

Treatments Design with Pretest) olarak belirlenmiştir (Barlow & Hayes, 1979). Bu desende araştırmanın yapıldığı üç grupta da müdahalede bulunulmuştur.

Çalışmanın yapıldığı grupların ilkinde dersler ÇS soru tipi(G1), ikincisinde AU soru tipi(G2) ve üçüncüsünde ise KT soru tipi(G3) üzerinden hazırlanan ders materyalleri ve bu materyallere uygun ev ödevleri dikkate alınarak gerçekleştirilmiştir. Bu doğrultuda gerçekleştirilen araştırmada derslerin işlendiği soru tipi ve ev ödevlerinin ÇS, AU ve KT tipi sorularla gerçekleştirilmiş grupların AU ve ÇS iki farklı madde tipi içeren ölçeklere göre başarı düzeylerinin farklı olup olmadığı belirlenmeye çalışılmıştır.

Çizelge 1'de görüldüğü gibi üç gruba AUT ve ÇST olmak üzere çalışma başında ön test olarak uygulanmıştır. G1'e ÇS soru tipi, G2'ye AU soru tipi, G3'e KT soru tipi ve bu soru tiplerine paralel ödevlendirmeler yapılmıştır. Çalışmanın sonunda da her üç gruba AUT ve ÇST son test olarak uygulanmıştır.

Çalışma Grubu

Çalışma araştırmacı öğretmenin göreve yaptığı okulda yürütülmüştür. Araştırmacı öğretmenin öğrencileri, okuldaki diğer çalışma arkadaşlarını ve okul yönetimini tanıması okul seçiminde etkili olmuştur. Aynı zamanda okulun fiziki yapısının yeterli olması (Etkileşimli tahtanın olması, beş yıllık bir binasının olması, fotokopi makinası gibi araç gereçler yönünden zengin olması) bu okulun tercih edilme sebebidir. Çalışmanın yürütüldüğü üç grubun matematik dersleri de araştırmacı öğretmen tarafından yürütülmüştür. Bu sayede öğretmen farklılığından kaynaklı olası etkiler kontrol altına alınmıştır. Araştırmanın örneklemi, 2019–2020 eğitim-öğretim döneminde Güney Doğu Anadolu Bölgesinin orta büyüklükte bir ilinin düşük sosyo-ekonomik seviyedeki öğrencilerin tercih ettiği bir imam hatip ortaokulunda 7/A(27) , 7/C(27) ve 7/D(23) şubelerine devam eden toplam 78 öğrenci oluşturmaktadır. A şubesi G1, C şubesi G2 ve D şubesi G3 olarak isimlendirilmiştir. Araştırma grubu, 7. sınıf şubeleri arasından belirlenen üç şubenin gruplara atanması rastgele seçilerek gerçekleştirilmiştir.

İşlem

Araştırma 7. Sınıf Matematik öğretim programı kapsamındaki, cebir öğrenme alanının alt öğrenme alanı olan; cebirsel ifadeler, eşitlik ve denklemler ile ilgili ünitelerinin konuları temel alınarak gerçekleştirilmiştir. G1 grubunda dersler ÇS soru tiplerine, G2 grubunda AU ve G3 grubunda KT soru tiplerine göre hazırlanmış etkinlikler ve ev ödevleri temel alınarak gerçekleştirilmiştir. G3 grubunda verilen ders materyalleri ayrıca hazırlanmamış, G1 ve G2 gruplarında kullanılan bir takım materyaller harmanlanarak ders işlenmiştir. Ayrıca ev ödevlerinde de G3 grubu, diğer iki gruptaki soru miktarlarının yarısı ile ödevlendirilmiştir. Bu şekilde grupların karşılaştığı soru miktarının eşitlenmesi amaçlanmıştır. Her üç grupta da öğrencilere verilen ev ödevleri haftalık 1 ders saati ayrılarak okulda öğrencilerle birlikte kontrol edilerek doğru cevaplar üzerinde tartışma yürütülmüştür. Bu yöntemle öğrencilerin soru tiplerine bağlı olarak öğrenmeleri hedeflenmiştir. Her üç grupta da öğrencilerin ders dışında aldıkları kurslarda cebir konusunun işlenmeyip, önceki konuların tekrar edilmesi sağlanmıştır. Bu şekilde öğrenciler sadece okulda gerçekleştirilen yöntemle ve miktarda cebir konusuna maruz bırakılmıştır. Aynı şekilde çalışma grubundaki hiçbir öğrenci özel ders almamıştır. Çalışmanın başında ve sonunda her bir gruba 16 maddeli ÇST ve 8 maddeli AUT uygulanmıştır. Aşağıda her üç grubun ders etkinliklerinden birer örnek verilmiştir. Çalışma 22 ders saati sürmüştür. Her üç grupta kullanılan ders materyalleri Resim 1'de gösterilmektedir.

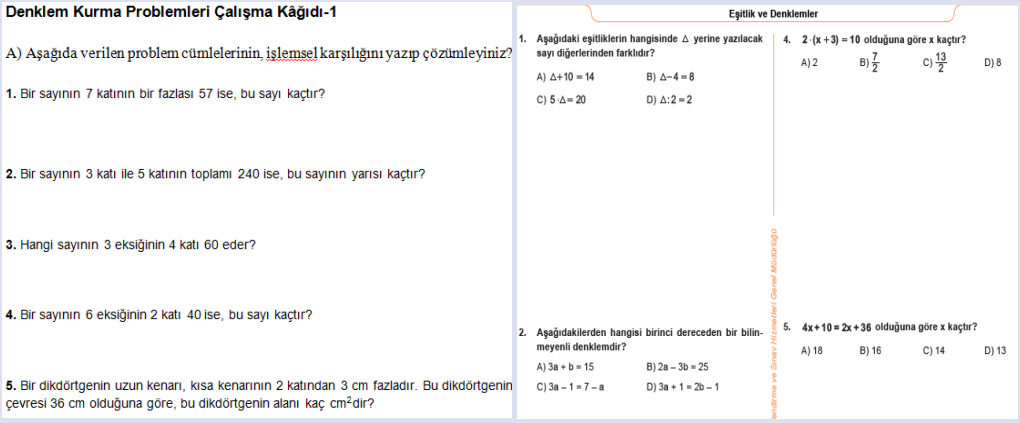
Veri Toplama Araçları ve Verilerin Toplanması

Veri Toplama Araçları

Bu çalışmada ÇS ve AU soru tiplerinden oluşmak üzere iki başarı ölçeği araştırmacılar tarafından geliştirilmiştir.

Çizelge 1. Araştırma modelinin simgesel görünümü

Gruplar	Başarı Testi (Ön-Test)	Denel İşlem	Başarı Testi (Son-Test)
G_1	AU ve ÇS	Çoktan Seçmeli soru tipi içeren materyaller ev ödevleri	AU ve ÇS
G_2	AU ve ÇS	Açık Uçlu soru tipi içeren materyaller ev ödevleri	AU ve ÇS
G_3	AU ve ÇS	Karma ev ödevleri	AU ve ÇS



Denklem Kurma Problemleri Çalışma Kâğıdı-1

A) Aşağıda verilen problem cümlelerinin, işlemsel karşılığını yazıp çözümleyiniz?

1. Bir sayının 7 katının bir fazlası 57 ise, bu sayı kaçtır?

2. Bir sayının 3 katı ile 5 katının toplamı 240 ise, bu sayının yarısı kaçtır?

3. Hangi sayının 3 eksiğinin 4 katı 60 eder?

4. Bir sayının 6 eksiğinin 2 katı 40 ise, bu sayı kaçtır?

5. Bir dikdörtgenin uzun kenarı, kısa kenarının 2 katından 3 cm fazladır. Bu dikdörtgenin çevresi 36 cm olduğuna göre, bu dikdörtgenin alanı kaç cm^2dir?

Eşitlik ve Denklemler

1. **Aşağıdaki eşitliklerin hangisinde Δ yerine yazılacak sayı diğerlerinden farklıdır?**

A) $\Delta + 10 = 14$ B) $\Delta - 4 = 8$

C) $5 \cdot \Delta = 20$ D) $\Delta : 2 = 2$

2. **Aşağıdakilerden hangisi birinci dereceden bir bilinmeyenli denklemdir?**

A) $3a + b = 15$ B) $2a - 3b = 25$

C) $3a - 1 = 7 - a$ D) $3a + 1 = 2b - 1$

4. **$2 \cdot (x + 3) = 10$ olduğuna göre x kaçtır?**

A) 2 B) $\frac{7}{2}$ C) $\frac{13}{2}$ D) 8

5. **$4x + 10 = 2x + 36$ olduğuna göre x kaçtır?**

A) 18 B) 16 C) 14 D) 13

Resim 1. Grupların kullandıkları ders materyalleri örnekleri

1. Çoktan seçmeli başarı testinin hazırlanma süreci

Veri toplama aracı şeklinde kullanılan matematik başarı testi, bazı özel eğitim kurumlarının yayınlarından, cebir ile ilgili lisansüstü tezlerden, geçmiş dönemlerde yapılmış Seviye Belirleme Sınavı (SBS), Temel Eğitimden Orta Öğretime Geçiş Sınavı(TEOG) ve Devlet Parasız Yatılı ve Bursluluk Sınavı (DPYB) soruları kullanılarak araştırmacılar aracılığıyla geliştirilmiştir. 7. sınıf cebir öğrenme alanının alt öğrenme alanlarını kapsayacak ev ödevleri verildiği için bu ödevlerin kazanımları ile örtüşecek ölçme aracı geliştirilmesi amaçlanmıştır. Hazırlanan Cebir Başarı Test'i(CBT) hedef-kazanım ilişkisini içerecek şekilde belirtke tablosu oluşturulmuştur. Oluşturulan başarı testinin madde analizi için 4 farklı okulda 237 8. Sınıf öğrencisi üzerinde uygulanmıştır. Ön uygulama için madde analizlerinin azami sınır olan 200 kişiden oluşacak bir grup üzerinde yapılması gerekmektedir(Akt.; Eren,2015; Crocker ve Algina, 2008). Belirtke tablosu ve ölçülecek kazanımın seviyesi madde tipinin eşleştirilmesinin nasıl yapılacağı ile ilgili karar almamızda yardımcı olmuştur. Ölçülecek davranışların sayısı ile test madde sayısı uyumlu olmasına dikkat edilmiştir. Testte bulunan madde sayısının dağılımı belirtke tablosunda bulunan kazanım-hedef seviye ilişkisi göz önünde bulundurularak kararlaştırılmıştır. 7. sınıf İlköğretim Matematik Öğretim P "Cebirsel" öğrenme alanı göz önünde bulundurularak rogramı kazanımlar belirlenmiştir (MEB, 2005). Belirtke tablosunda belirlenen kazanımları ölçmek amacıyla 32 maddeden oluşan ölçek hazırlanmıştır. Hazırlanan sorular, dil, anlatım; kapsamı ve öğrenci düzeyine uygunluğu bakımından mesleki deneyimleri 4- 15 yıl arasında olan 4 tane matematik öğretmenine ve 3 tane matematik eğitimi alan uzmanı görüşlerine sunulmuştur. Deneyimli eğitimciler ve matematik eğitimi alan uzmanlarının önerileri ve yönlendirmeleri doğrultusunda gerekli düzeltme ve değişiklikler yapılmış ve 8 adet soru ölçekten çıkarılmıştır. Ölçek 24 madde olarak kararlaştırılmıştır. Çalışmanın yapıldığı il merkezindeki araştırmanın yapıldığı okul dahil, 3 farklı ortaokulda 8. Sınıf 253 öğrenci üzerinde yapılmış; ancak 23 tane ölçek sonucunun öğrenciler tarafından ciddiyetle çözülmediği gözlenmiştir. Dolayısıyla 230 adet ölçek üzerinden madde analizi yapılmıştır. Ölçeğin pilot uygulama verilerinin madde analizi yapılırken, madde güçlük ve ayırt edicilik indeksleri bulunmuştur. Ayırt edicilik indeksi 0.30'nin altında olan 5 madde(1,4,5,6,8. maddeler) testten çıkarılmıştır. Bu maddelerin testten çıkarılması kapsam geçerliliğini zedelemediği gözlenmiştir. Güçlülük indeksi 0.70 değerini aşan 3 soru testten çıkarılmıştır. Benzer alt kazanımları ölçtüğü tespit edilen 3(13,17,21. maddeler) madde ise testin kullanışlılığını güçlendirmek amacıyla idealden uzak olmalarına ve güçlülük indeksi 0.70'in üzerinde oldukları için testten çıkarılmıştır. Pilot uygulama öncesinde, başarı testinin madde sayısına karar verilirken, uygulamayı takiben testten çıkarılacak soruların testin kapsam geçerliliğini olumsuz yönde etkilememesi amacıyla soru miktarı fazla tutulmuştu. Pilot uygulamayı takip eden süreçte ölçeğin nihayi halinin kapsam geçerliliğinin sağlandığı, belirtke tablosuna bakılarak karar verilmiştir. Cronbach Alpha katsayısı, ölçeğin maddeler arasındaki iç tutarlığı belirlemek niyetiyle kullanılmaktadır. SPSS 20.0 programı kullanılarak gerçekleştirilen güvenirlilik analizi neticesinde başarı testinin ilk formunun Alpha katsayısı 0.855 olarak bulunmuş ve bu değer ile testin güvenilir olduğu kanaatine varılmıştır. Madde toplam korelasyon katsayısına bakıldığında 0.30'un altında olan maddeler testten çıkarılmalı 0.30 - 0.20 aralığında bulunan sorular mümkünse ölçekte kullanılmamalı 0.20'nin altında kalan sorular ise ölçekten ayıklanmalıdır (Büyüköztürk, 2010). Madde toplam korelasyon değeri 0.30 değerine ulaşamayan 4. ve 13. maddeler olmak üzere 2 madde testten çıkarılmıştır. Sonuç olarak pilot uygulama sonrası testte bırakılan 16 maddenin tekrar madde ayırt edicilik ve güçlülük analizi yapılarak analiz sonuçları tablo olarak dizayn edilip madde analiz tablosunda verilmiştir. Başarı testinin nihayi halinin SPSS 20.0 programı kullanılarak

yapılan güvenirlilik analizi sonucu Alpha katsayısı 0.874 değerinde olduğu görülmüştür. Bu değer ölçeğin güvenirliliğine karar vermek için gerekli kriteri sağladığı sonucuna ulaşılmıştır. Testin puanlamasında, hazırlanan testte her bir maddenin doğru cevabı 2,5 puandır, hatalı veya yanıtlanmayan sorulara 0 puan verilmiştir. Test 16 soru içerdiği için, tüm maddelerin doğru cevaplandırılması durumunda en yüksek puanın 40 olacağına karar verilmiştir. Öğrenci kağıtlarının değerlendirilmesinde hatalı cevapların doğru cevap verilen maddelerin puanını etkilememektedir.

2. Açık Uçlu Başarı testi hazırlama süreci

Bu çalışmada 2. veri toplama aracı, AU Cebir Başarı Testi, 8 açık uçlu maddeden oluşturulmuş olup, araştırmacılar tarafından geliştirimiştir. MEB açık kaynakları ve 7. sınıf test kitaplarından yararlanılarak soru havuzu oluşturulmuştur. İlk aşamada 12 sorudan oluşan test, 3 alan eğitimi uzmanı ve 4 matematik öğretmeninin görüşleri doğrultusunda kapsam geçerliliğini bozmayacak şekilde 4 maddesi çıkarılıp, kalan 4 maddede de kısmi düzenlemeler yapıldıktan sonra açık uçlu 8 madde olarak belirlenen teste son hâli verilmiştir. Aynı okulda bir üst kademede bulunan öğrenciler üzerinde pilot uygulaması yapıldıktan sonra gerekli düzeltmeler yapılıp teste son hali verilmiştir. AU testin puanlaması, değerlendirme rubriği ile yapılmıştır. Puanlama 0-3 arasında olacak şekilde yapılmıştır. Soruyu tam cevabı 3 ile puanlanırken sorunun cevabını boş bırakan veya doğru çözüm ile ilişkili olmayan yanıtlar 0 ile puanlanmış, kısmi cevaplar ise açıklama ve çözümü tam puana yakın ise 2 ile değilse 1 ile puanlanmıştır.

Çalışmanın uygulama süreci, 22 (4,5 hafta) ders saati sürmüştür. AU ve ÇS başarı testleri, bütün gruplara çalışmanın başında, ön-test çalışmanın sonunda ise son-test olarak iki ders saatinde uygulanmış olup, bu 4 saatlik sınav saati 22 saatlik uygulama sürecinin dışında tutulmuştur.

Veri Analizi

Analiz sürecinin başında verilerin normal dağılıma uyup uymadığı kontrol edilmiştir. Bunun için çarpıklık ve basıklık değerlerine bakılmıştır (Tablo 3). Çarpıklık ve basıklık katsayılarının -2 ile +2 arasında çıktığı görülmüş olup verilerin normal dağılım gösterdiği sonucuna varılmıştı ve parametrik testlerin kullanılmasına karar verilmiştir.

Gruplardaki öğrencilerin başarılarının hem AU hem de ÇS testlerden elde edilen ilk ve son test puanları, bağımlı *t*-testi ile analiz edilmiştir. Aynı zamanda her üç grubun ön test puanları yine iki farklı ölçekte çalışmanın başında Anova testi ile karşılaştırılmıştır. Çalışmanın sonunda ise grupların başarı düzeylerinde fark oluşup oluşmadığı hem AU hem de ÇS test puanları Anova testi ile analiz edilmiştir.

Çalışmada grupların normal dağılıma sahip olmaları Çarpıklık ve Basıklık değerleri göz önünde bulundurularak kararlaştırılmıştır. Çarpıklık değeri -.189 ile 1.788 aralığında, basıklık değeri ise -1.462 ile 1.945 aralığında değiştiği gözlenmiştir. Normal dağılım varsayımını sağlaması için, çarpıklık ve basıklık değerlerinin -2.0 ile +2.0 aralığında olması yeterli görülmektedir(Tabachnick and Fidell, 2013).

Verilerin Toplanması

Çizelge 2. Madde güçlülük ve ayırt edicilik indeksleri

Madde No	Madde Güçlülüğü(pi)	Madde Ayırt Ediciliği(rj)	Madde No	Madde Güçlülüğü(pi)	Madde Ayırt Ediciliği(rj)
1	0,86	0,35	9	0,77	0,51
2	0,71	0,51	10	0,63	0,47
3	0,79	0,41	11	0,71	0,67
4	0,72	0,59	12	0,67	0,74
5	0,76	0,50	13	0,74	0,59
6	0,41	0,38	14	0,74	0,68
7	0,68	0,44	15	0,75	0,54
8	0,59	0,78	16	0,60	0,73

Çizelge 3. Basıklık ve çarpıklık katsayıları

		Ön-Test		Son-Test	
		Çarpıklık	Basıklık	Çarpıklık	Basıklık
Çoktan Seçmeli	G_1	-,182	-1,267	,466	,012
	G_2	1,081	,619	-,067	-1,462
	G_3	1,118	,848	,365	-,789
Açık uçlu	G_1	1,241	1,632	,401	-,597
	G_2	,590	-,242	-,189	-,599
	G_3	1,788	1,945	,930	-,077

Bulgular

Araştırmanın bulguları iki bölümde değerlendirilmiştir. Birinci bölümde, hem ÇS hem de AU test için, her bir grubun kendi içinde bağımlı gruplar *t* testi kullanılarak ön-test ve son-test başarı puanları farkı hesaplanmıştır. Üç grubun da çalışmanın başındaki AU ve ÇS başarı ön-test puanları ile çalışmanın sonundaki son test puanları arasında anlamlı farklılık olup olmadığı belirlenmiştir. Grupların kendi içlerinde etki değerleri hesaplanarak 3 grup kendi aralarında etki değerlerine göre kıyaslanmıştır.

İkinci bölümde üç grubun ön test puanları dikkate alınarak son test puanları karşılıklı kıyaslanması için, belirli varsayımları karşılanması durumunda Anova karşılanmaması durumunda ise Kruskal Wallis testinin yapılmasına karar verilmiştir.

Birinci bölüm de her bir grubun hem ÇS hem de AU test puanları dikkate alınarak çalışmanın sonunda başarılarında anlamlı bir farkın olup olmadığı araştırmaya cevap verecek test, normallik varsayımını sağlaması durumunda Bağımlı örnekler *t* testi; sağlayamadığı durumda ise bu testin nonparametrik testi olan wilcoxon işaretli sıralar testi uygulanacaktır.

Üç grubun öğrencilerinin başarı testi, ön-test ve son-test puanları farkının anlamlı olup olmadığını incelemek amacıyla bağımlı örneklem *t*-testi ile değerlendirilmesine karar verilmiştir. Ancak bu testin yapılabilmesi için bazı varsayımların sağlanıp sağlanmadığı kontrol edilmelidir. Bu varsayımlar; dağılımın normal dağılım olması ve değişkenlerin en az aralık ölçeğinde olması şartlarıdır. Bu çalışma başarı testi olmasından dolayı aralık ölçeğindedir. Ancak verilerin normallik kriterlerini sağladığının kontrol edilmesi amacıyla her bir grubun fark puanlarının normallik testi yapılmıştır. Çarpklık ve basıklık katsayılarına bakılarak da normallik koşulunu sağlayıp sağlamadığına karar verilmiştir. Bu üç grubunda çarpıklık basıklık katsayılarının -2 ile +2 aralığında bulunduğu Tablo 3'ten görülmektedir. Bu durumda üç grubun ön ve son test puanlarına bakılarak bağımlı örnekler *t* testi yapılabileceği kararına varılmıştır. Bağımlı örnekler *t* testi sonucu Çizelge 4'te gösterilmiştir.

Çizelge 4. Her üç gruptaki öğrencilerin AU ve ÇS sorular ön test ve son testinden aldıkları puanlar için bağımlı gruplar t-Testi sonuçları

	Grup	Ölçek sırası	N	$\bar{X}$	Ss	t	sd	p	Etki Değeri
Çoktan Seçmeli Sorular	G_1	Ön-Test	27	11.15	4.63	-6.07	26	.000*	.405
		Son-Test	27	20.09	7.61				
	G_2	Ön-Test	28	12.23	5.53	-6.76	27	.000*	.478
		Son-Test	28	24.28	9,58				
	G_3	Ön-Test	23	12.28	4,50	-3.98	22	.001*	
		Son-Test	23	20.43	9,51				.235
Açık Uçlu sorular	G_1	Ön-Test	27	7,96	7,93	-6.47	26	.000*	.447
		Son-Test	27	19,19	11,19				
	G_2	Ön-Test	28	7,75	4,81	-7.96	27	,000*	.610
		Son-Test	28	20,96	9,56				
	G_3	Ön-Test	23	4,34	4,94	-5.91	22	,000*	.427
		Son-Test	23	13,60	9.92				

*p<.05

Çizelge 4 incelendiğinde her üç grupta da deneysel işlem öncesi ve sonrasındaki ÇS test puanları arasındaki farkın anlamlı bulunduğu görülmektedir [G1 için *t*(26)= -6.07; G2 için *t*(27)= -6.76; G3 için *t*(22)= -3.98]. Etki değerleri için Cohen (1988)'in formülü kullanılarak, [G1 için η^2 = 405; G2 için η2 = 478; G3 için η^2 = 235] olarak hesaplanmıştır.

Benzer şekilde tüm gruplar deneysel işlemden önceki ve sonraki AU test puanları arasında anlamlı fark bulunmuştur [G1 için t(26)= -6.469; G2 için t(27)= -8.305; G3 için t(22)= -6.057]. Etki değerleri gözden geçirildiğinde [G1 için η2 = 447; G2 için η2 = 610; G3 için η2 = 427] olarak hesaplanmıştır.

Her bir grup için çoktan seçmeli ve açık uçlu testlerde gözlenen anlamlı farka ait etki büyüklükleri karşılaştırıldığında Cohen (1988) etki değerinin sınıflandırılmasında; .01, küçük, .06 orta, .14 büyük etki değeri olarak belirlenmiştir. Bu durumda tüm gruplarda büyük etki değerine denk gelen bir ilerleme gözlenmiştir. G2 grubu AU testte en büyük gelişmeyi gösterirken, G3 grubu ÇS testte diğer gruplar ile karşılaştırıldığında en düşük etki değerine ulaştığı görülmektedir. İkinci bölümde ise üç grubun ön test puanları dikkate alınarak son test puanların karşılıklı kıyaslanması için, belirli varsayımların karşılanması durumunda Anova, karşılanmaması durumunda ise Kruskal Wallis testinin yapılması planlanmıştır. Anova testinin uygulanması için her bir grubun diğerinden bağımsız olması, her bir grubun verilerinin normal dağılıma sahip olması ve ortalamaları kıyaslanacak grupların varyanslarının eşit olması gerekmektedir.

Çizelge 5. Grupların ÇS ve AU ön test puanları için yapılan Tek Yönlü Anova

	Varyansları kaynağı	Kareler toplamı	sd	Kareler ortalaması	F	p
Çoktan seçmeli	Gruplar arasında	21,656	2	10,828	,461	,632
	Gruplar içinde	1761,062	75	23,481		
	Toplam	1782,718	77			
Açık uçlu	Gruplar arasında	200,057	2	100,028	2,682	
	Gruplar içinde	2797,430	75	37,299		,075
	Toplam	2997,487	77			

Çizelge 6. Grupların ÇS ve AU son test puanları için yapılan Tek Yönlü Anova

	Varyansları kaynağı	Kareler toplamı	sd	Kareler ortalaması	F	p	Etki değeri
Çoktan seçmeli	Gruplar arasında	293,785	2	146,892	1,863	,162	
	Gruplar içinde	5912,385	75	78,832			
	Toplam	6206,170	77				
Açık uçlu	Gruplar arasında	724,971	2	468,160	3,446		
	Gruplar içinde	7890,517	75	107,274		,037	,084
	Toplam	8615,487	77				

Çizelge 7. Gruplarının AU son test puanları için yapılan Scheffe Testi sonuçları

İlişkili Gruplar		Ortalamalar arasındaki fark	P
G_1	G_2	-1,779	,814
	G_3	5,576	,167
G_2	G_1	1,779	,814
	G_3	7,355	,044
G_3	G_1	-5,576	,167
	G_2	-7,355	,044

Çizelge 5'in ÇS ve AU ön test sonuçları incelendiğinde; çalışma öncesi uygulanan Başarı Testlerinin verileri analiz edilmesi ile grupların başarı düzeyleri arasında bulunan farkın anlamlı düzeyde olmadığı görülmektedir (p>0,05). Dolayısıyla, uygulama öncesi kontrol ve deney gruplarının ön bilgilerinin birbirine yakın olduğu kabul edilmiştir.

Deney gruplarının ön test sonuçları karşılaştırıldığında farkın anlamlı olup olmadığını incelemek amacıyla SPSS programında One-Way ANOVA uygulanmıştır. Tablo'5'te verilen One-Way ANOVA sonucu incelendiğinde ÇS sorular için; F değerinin 0.632 ve p = 0.632 > 0.05 olarak saptandığı; AU sorular için; F değerinin 2,682 ve p = 0.075 > 0.05 olarak saptandığı görülmektedir. Bu değerler 0.05 düzeyinde anlamlı değildir. Diğer bir ifadeyle, gruplar arasında deneysel işlem öncesinde ÇS ve AU test sonuçlarında anlamlı bir farklılık olmadığından grupların denk olduğu söylenebilir.

Çizelge 6'da görülmektedir ki, deneysel uygulama sonrası uygulanan ÇS ve AU testin verileri analiz edildiğinde grupların başarı düzeyleri arasında ÇS test sonuçları dikkate alındığında farkın anlamlı olmadığı görülmektedir (p<0,05), ancak; AU test sonuçlarına göre anlamlı bir farkın olduğu görülmektedir (p<0,05). Bu farkın 0,084 düzeyinde geniş etki büyüklüğünde olduğu söylenebilir (Green ve Salkind, 2005). Bu durumda AU testin son-test olarak uygulandığı grupların son-test başarıları karşılaştırıldığında farkın anlamlı düzeyde olduğu görülmüştür. AU son test sonuçlarını dikkate alarak Tek Yönlü Anova'dan gözlemlenen anlamlı farkın kaynağını tespit etmek amacıyla post hoc testlerden Scheffe testi yapılmış ve elde edilen sonuçları Çizelge'7 de verilmiştir.

Farklı soru tipleri üzerine ders alan öğrenci gruplarının AU son-test puanlarını ikili olarak karşılaştırmak niyetiyle tamamlayıcı testlerden (Post-Hoc) Scheffe testi yapılmıştır. Tablo 7'de belirtilen sonuçlara bakıldığında, G1 ile G2 ve G1 ile G3 arasında son test puanları kıyaslandığında farkın anlamlı olmadığı görülmektedir (p>0,05). Ancak, G2 ile G3 öğrencilerinin AU son-testi puanları karşılaştırıldığında farkın G2 lehine anlamlı olduğu görülmektedir (p<0,05). Bu durumda AU soru tipleri ile eğitim gören ve bu soru tipleri üzerine ev ödevi verilen G2 grubu, hem AU hem de ÇS soru tipleri ile eğitim gören ve bu soru tipleri üzerine ev ödevi verilen G3 grubundan AU son test puanı daha yüksek çıkmıştır.

Tartışma, Sonuç ve Öneriler

"Matematik dersinde ÇS, AU ve KT uygulanan her bir grubun AU ve ÇS test puanlarına göre ön-test ile son-test puanları arasında anlamlı fark var mıdır?" şekliyle ifade edilen araştırma problemine cevap olarak şunlar ifade edilebilir. Grupların ön-test ve son-test puanlarının karşılaştırılması ile elde edilen bulgular; öğrencilerin hem ÇST hem de AUT puanları karşılaştırıldığında üç grupta da farkın son test puanları lehine anlamlı olduğu görülmektedir. Etki değerlerine bakılarak gruplar karşılaştırıldığında hem AUT hem de ÇST puanları dikkate alınarak hesaplanan etki değerlerinde, en güçlü etkiyi yapan grubun AU soru tipleriyle öğrenim gören G2 grubu iken; en zayıf etki değerine sahip grubunda KT soru tipleriyle öğrenim gören G3 grubu olduğu görülmektedir. Her üç grupta da farklı yöntemlerle ders işlenmesine ve dersin işlendiği yaklaşıma göre yapılan ödevlendirmeye rağmen; en başarılı grubun AU sorularla ders işlenen ve ödev verilen grubun olması, bu grubun kavramsal öğrenmelerinin daha güçlü bir şekilde gerçekleşmelerinden kaynaklandığı düşünülmektedir. AU sorular ve çalışma yaprakları ile öğretimin kavramsal öğrenmeyi güçlendirdiği ifade edilmiştir (Baki, 2018). KT yöntemi ile öğretim etkinliklerinin uygulandığı G3 grubunda ise hem AU hem de ÇS test başarısının en düşük olması, öğrencilerin herhangi bir soru tipine yeterince odaklanamamalarından kaynaklandığı düşünülmektedir.

Grupların test türüne göre başarı durumları karşılaştırıldığında, tüm grupların hem AUT hem de ÇST başarılarında anlamlı bir şekilde artış görülmüştür. Ancak üç grubun da AUT başarısındaki artışın etki değeri daha yüksek olduğu görülmüştür. Yani her bir grubun AUT başarılarının etki değeri daha yüksek çıkmıştır. Bunun sebebi olarak AU ön test puanları çok düşük olduğundan son test puanlarındaki artış ÇST ile kıyaslandığında daha yüksek görünmektedir. Bu da AU test puanlarının etki değerlerinin daha yüksek çıkmasına neden olmuştur. Oysa ÇS maddelerden oluşan testte düzeltme işlemi yapılmadığı için şans faktörünün etkisiyle, ön test puanları AUT ön test puanından yüksek çıkmıştır. Yani ÇS maddelerden oluşan testteki başarının artışı daha düşük etki değerde olması ile sonuçlanmıştır. Gültekin ve Demirtaşlı(2012), TIMSS 2007 verilerini kullanarak yaptıkları çalışmalarında, çoktan seçmeli soruları içeren ölçeğin, açık uçlu soruları içeren ölçek ile karşılaştırıldığında, öğrencilerin matematik başarısı ile ilgili daha az bilgi verdiği neticesine ulaşılmıştır. Öksüz ve Demir, (2019) çalışmasında fen ve sosyal bilimler derslerinde öğrencilerin AU ve ÇST madde tiplerinde, Fen Bilimleri dersinde ÇST'nin AU ya göre daha etkili olduğunu, sosyal bilimler dersinde ise herhangi bir farklılığın olmadığı sonucuna ulaşıldığı görülmüştür. Buda ders türünün bir değişken olabileceğini göstermektedir.

İkinci araştırma sorusu olan "Matematik dersinde ÇS, AU ve KT uygulanan grupların, AU ve ÇS test puanları arasında anlamlı fark var mıdır?" cevap olarak şunlar ifade edilebilir; üç grubunda ANOVA son test başarı puanları karşılaştırıldığında ÇS son-test puanında anlamlı fark görülmemiştir. Ancak AU son test puanında ise anlamlı bir farklılık görülmüştür. Bu farklılığın AU madde tipi ile öğretim etkinliklerinin yapıldığı G2 ile hem AU hem de ÇS madde türlerinin uygulandığı G3 grubu arasında G2 lehine olduğu görülmektedir. Aydın ve Önder(2010)'da yaptıkları çalışmada ÇS ve AU soru tiplerinin başarı üzerinde etkisinin araştırmada AU soru tipleri ile eğitim gören öğrencilerin diğer soru tiplerinde de daha başarılı oldukları görülmüştür. Benzer sonuçlar Cambaz(2010) çalışmasında ilköğretim düzeyindeki öğrenciler ile yaptığında AU maddeler ile eğitimin yapıldığı grupta başarının anlamlı miktarda daha yüksek çıktığı görülmüştür. AU soru tiplerinin çözümü, öğrencinin kendi üretimi olan içselleştirildiği ve öğrenci tarafından yapılandırılan(Haladyna, 1997) öğrenme çıktısına yol açtığı için daha nitelikli bir öğrenme gerçekleşmektedir. KT içerikli soru tipleri ile eğitim yapılan grubun en düşük etki değerine sahip grup olması literatüre aksi bulgu olarak değerlendirilmektedir. Gültekin ve Çıkrıkçı-Demirtaşlı (2012) çalışmalarında KT'nin öğrencilerin başarılarını artırmada daha iyi hizmet ettiklerini belirtmişlerdir. Ölçme araçlarındaki çeşitliliğin bulunduğu öğrenmelerin farklı soru türleri ile ölçülmesi sonuçların daha sağlıklı olmasını sağlamaktadır (Buket, 2015). Benzer şekilde, Küçüker(2017) TIMSS 2011 Matematik başarı testi verilerini dikkate alınarak gerçekleştirilen çalışmasında, madde tiplerinin çeşidine göre, sorulara doğru yanıt verme yüzdelerinin ortalamaları hesaplanmıştır. Yapılan analizler neticesinde; öğrencilerin ÇST puanlarının AU test puanlarından daha yüksek bulunmuştur. Bu bulgu literatürdeki araştırmaların aksi yönündedir, dolayısıyla bu sonuca kaynaklık eden faktörler hakkında değerlendirme yapmak için konu ile ilgili çalışmaların artırılmasına ihtiyaç vardır. Yapılacak nitel çalışmalar yardımıyla elde edilecek zengin veri seti ile bu çıktının derinlikli çalışması yapılarak, sorunun sebepleri daha açık hale gelebilir.

Bu araştırmanın bulguları Tablo 6'da görüldüğü gibi ÇST puanları üzerinde ANOVA son test sonuçlarında anlamlı farklılık görülmemiştir. Bu bulgu sınava hazırlık sürecinde kullanılan soru tipinin ÇS test üzerinde belirleyici bir etkisinin olmadığını göstermektedir. Bu yönü ile Aydın ve Önder(2010) çalışması ile benzer sonuçlara ulaşılmıştır. Bunun sebebi olarak öğrencilerin uzun bir süredir ÇS madde tipindeki testlere maruz kalmaları bu alandaki deneyim ve becerilerinin kuvvetlenmiş olması beş haftalık bir eğitim sürecinde bu deneyimlerini etkileyecek yeterlilikte bir zaman aralığı olmaması olarak görülmektedir. Öksüz ve Demir (2019) çalışmalarında, öğrencilerin Fen bilgisi ve sosyal bilgiler derslerinde benzer soruların hem AU hem de ÇS madde tipine uyarlanmış hallerinde öğrencilerin ÇS test puanlarının AU test puanlarından yüksek olduğu görülmüştür. İlhan vd., (2020) çalışmalarında üst düzey sınıfların ÇST puanlarında, öğrenci başarısının alt düzey sınıflardan daha yüksek çıktığı, AUT maddelerinde ise tersi durumun olduğu bulgusuna ulaşılmıştır. Bu araştırmada, üst düzey sınıf seviyesinde olan yedinci sınıflarda ÇS madde türüne öğrencilerin aşinalığı ve deneyimlerinin kuvvetli olması bu

madde türünde başarının kararlılık kazanmasına ve beş haftalık bir öğretim etkinliğinin etkisinin zayıf olabileceği olarak yorumlanmaktadır. Bu çalışma matematik dersindeki belli kazanımlarla sınırlı olsa da; matematiğin diğer kazanımları ve farklı dersler içinde geçerli olabilecek bulgular barındırdığı değerlendirilmektedir.

Sonuç olarak, AU, ÇS ve KT madde tipini içeren öğretim etkinlikleri ve ev ödevlerinin öğrencinin başarısını AUT ve ÇST ile karşılaştırıldığında AU soru tipleri ile eğitimin yapıldığı G2 grubunun en yüksek başarıyı gösterdiği görülmektedir. Çalışmanın yapıldığı üç grup karşılaştırıldığında G2 grubunun sadece G3 grubundan anlamlı bir farkının olduğu diğer gruplar arasında benzer seviyede bir gelişim gözlenmiştir. Bu çalışmada matematik dersi üzerine odaklanılmıştır, diğer disiplinler üzerinde de benzer çalışmalar yapılabilir. Yapılan çalışmanın benzeri merkezi sınavların yapıldığı sınıf seviyesi veya bu sınıf seviyesine yakın sınıflar ile bu sınıf seviyelerine uzak seviyedeki, farklı iki sınıf seviyesine uygulanıp sonuçlarının karşılaştırılması önerilmektedir. Yapılacak söz konusu çalışmalar ile farklı sınıf düzeyindeki öğrencilerin karşılaşacakları muhtemel soru tipi ile öğrenme yöntemi eğiliminin zamana bağlı olarak nasıl değiştiği gözlemlenebilecektir. Eğitim etkinliklerinin gerçekleştirildiği madde tipinin kavram yanılgıları üzerinde etkili olmaktadır (Aydın ve Delice, 2010; Demirbilek, 2015). Bu çalışmada, AU maddeler ile öğretim etkinliklerinin gerçekleştiği grup, ÇS maddeler ile öğretim etkinliklerinin gerçekleştirildiği grup ile karşılaştırıldığında cebir ile ilgili kavram yanılgılarının daha az olması beklenir. İleride yapılacak benzer çalışmalarda kavram yanılgılarının da incelenmesi önerilmektedir.

Bu çalışmanın yapıldığı okul imam hatip ortaokulu olmasından kaynaklı, gruplar cinsiyete bağlı ayrışmış bulunmaktaydı. Yanı G1'de sadece erkekler varken diğer iki grup kız öğrencilerden oluşmaktadır. Bu durum çalışmanın sınırlılıkları arasında kabul edilmiştir. Lyons-Thomas vd. (2014) yaptıkları çalışmada erkeklerin matematikte kızlara göre daha başarılı olmakla birlikte, ÇS soru tiplerinde de erkekler daha başarılı iken kız öğrencilerinde AU soru tiplerinde daha başarılı oldukları saptanmıştır. Cinsiyet yönünden karma sınıfların olduğu okullarda bu çalışmaya benzer çalışmalar yapılıp, kız ve erkek öğrenci değişkenine göre yapılmasına odaklanılabilir.

Bu çalışmada nicel veriler ile AU soruların öğrenci başarısına etkisi ÇS testlerden daha yüksek başarıyı sağladığı sonucuna ulaşılmışsa da, öğrencilerle yapılacak klinik mülakatlar ile bu sonucun nedeni derinlemesine ortaya konabilir. Bu bağlamda çalışma yapacak araştırmacıların bunu dikkate alarak araştırma yapmalarının daha etkili olacağı öngörülmektedir.

Eğitim politikalarını belirleyen yöneticiler, bu çalışmada elde edilen sonuçların eğitim sistemine entegre edilmesini ve okullarda AU soru tiplerine ağırlık verilmesinin MEB (2023) tarafından ifade edilmiş olmak ile birlikte AU soru tiplerinin yaygınlaştırılarak merkezi sınavlarda sorulmasının alt yapısı kurulmalıdır.

Extended Abstract

Introduction

The curriculum consists of four basic elements that are in a dynamic relationship with each other: objectives, content, learning-teaching process and assessment (Arslan et al., 2018; Wolf et al., 2007). At the end of the educational process, the level of achievement of the outcomes determined in the program is determined by evaluation studies (Baykul, 2010; Van den Akker, 2013). In this sense, the evaluation element provides feedback to the stakeholders of education and gives the educational process a dynamic (Turgut & Baykal, 2014; Baykul, 2010; Öksüz & Demir, 2019) and cyclical (National Council of Teachers of Mathematics [NCTM], 1995) structure. Assessment is also the most important component that increases the quality of education (Gültekin, 2011; National Resarcher Concil [NRC], 2001).

The measurement tools used in the measurement and evaluation process vary as multiple-choice, complex multiple-choice, open-ended, semi-structured and short-answer questions (MoNE, 2015). Which one to prefer between OER examinations and MLT should be decided in line with the needs of the exam practitioners. The advantages of AU question types are that their items are easy to write, they have a structure that can measure students' high-level cognitive skills (Herman et al., 1994), they are easy to prepare, and they allow the teacher to evaluate various techniques used by students (Chaoui, 2011). The disadvantages are thought to include the concern that the scoring may not be fair. Similarly, the advantages of ES question types can be listed as ease of scoring, testing a large number of people in a single session, easy and objective scoring of responses (Bible et al., 2008), and relatively high content validity of the test (Norman, 1995; Demirtaşlı, 2010; Özçelik, 2011). The disadvantages are that it is difficult to write quality questions (Demirtaşlı, 2010) and a large sample is needed for reliability analysis in the preparation of ES tests.

Looking at the studies examining the effect of question type on student achievement, Eren (2015) examined the relationship between the scores of two measurement tools measuring ES and CT item types, as well as the opinions of teachers and students about these item types, and it was observed that both teachers and students preferred the ES test format more. Similar results were obtained by Demir (2010) who found that students in Turkey preferred MC question types to AS question types in mathematics tests and were more successful. Özkan and Özarslan (2018) examined student responses according to the question types of the PISA 2003-2012 exam, and found that while students showed higher success in ES question types, they showed lower success in Complex Multiple Choice and AU question types than in ES question types.

Toker and Green (2016) conducted a study on TIMSS 2011 data and found that ES item types were easier for Turkish students. It was found that there were AU item types that lowered the overall scores of the students.

In exams such as PISA and TIMSS, it has been observed that the reason why the achievement in AUT question types is lower than the achievement in ES question types is that Turkish students are not accustomed to item types other than ES items used in these exams (Demir, 2010), and students' performance is higher in question types that they are experienced and familiar with (Ben-Simon and Cohen 2004).

The general purpose of this study is to investigate the effect of homework assignments and instructional activities given to students according to question types on students' mathematics achievement according to CT, AUT and ES.

Method

The research design used in this study is called "Alternative-Treatments Design with Pretest" (Barlow & Hayes, 1979). In this design, interventions were made in all three groups in which the research was conducted.

In the first group, the lessons were carried out by taking into account the course materials prepared on the ES question type (I1), in the second group on the AU question type (I2) and in the third group on the CT question type (I3) and homework assignments appropriate to these materials. In the research conducted in this direction, the researchers explored whether the achievement levels of the groups in which the lessons were taught and homework assignments were carried out with ES, AS and CT type questions were different according to the scales containing two different item types of AS and ES.

Results, Discussion, Pedagogical Implications

The general purpose of this study is to investigate the effects of homework assignments and instructional activities on students' mathematics achievement according to CT, AU and ES in middle school mathematics course. The first problem investigated in this study is "Is there a significant difference between the pre-test and post-test scores of each group in which ES, AU and CT were applied in the mathematics lesson according to the AU and ES test scores?" The findings obtained by comparing the pre-test and post-test scores of the groups are as follows; when both the ES and AUT scores of the students were compared, a significant difference was found in favor of the post-test scores in all three groups. When the groups were compared by looking at the effect values, it was seen that the group with the strongest effect is the I2 group with AU question types, while the group with the weakest effect value was the I3 group with CT question types.

As a result, it is seen that the I2 group, in which the instructional activities and homework assignments including AU, ES and CT item types were used to improve the student's achievement compared to AUT and EST, showed the highest achievement. When the three groups in which the study was conducted were compared, it was observed that the G2 group had a significant difference only from the G3 group, and a similar level of improvement was observed among the other groups. This study focused on mathematics, similar studies can be conducted on other disciplines. It is suggested that a similar study should be applied to the grade level in which central exams are held or to the grades close to this grade level and to two different grade levels far from these grade levels and the results should be compared.

Etik Kurul İzni

Bu çalışmanın Adıyaman Üniversitesi Sosyal ve Beşeri Bilimler Etik Kurul Komisyonundan 27.02.2024 tarihli 22/03/2024-31 sayılı toplantısı kararıyla etik kurul izni, E-72214596-604.02.98468526 sayılı izin belgesi ile MEB uygulama izni alınmıştır.

Araştırmanın Etik Taahhüt Metni

Yapılan bu çalışmada bilimsel, etik ve alıntı kurallarına uyulduğu; toplanan veriler üzerinde herhangi bir tahrifatın yapılmadığı, karşılaşılacak tüm etik ihlallerde “Cumhuriyet Uluslararası Eğitim Dergisi ve Editörünün” hiçbir sorumluluğunun olmadığı, tüm sorumluluğun Sorumlu Yazara ait olduğu ve bu çalışmanın herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiş olduğu sorumlu yazar tarafından taahhüt edilmiştir.

Kaynaklar

Ak, H. (2002). *Okul öncesi öğretmen adaylarının yaratıcılık ile ilgili metaforları*. [Poster bildiri]. 6. Uluslararası Eğitim Araştırmaları Kongresi, Ankara.

Anderson, R. S. (1998). Why talk about different ways to grade? The shift from traditional assessment to alternative assessment. *New Directions for Teaching and Learning*, (74), 5–16. https://doi.org/10.1002/tl.7401.

Arslan, A., & Eker, C.(2018). “Öğretim İlke ve Yöntemleri”. Nobel

Aydın, E., & Delice, A. (2008). Ölçme ve değerlendirmeye kavram yanılgıları perspektifinden bir bakış. MF Özmantar, E. Bingölbali & H. Akkoç (Edt.). Matematiksel kavram yanılgıları ve çözüm önerileri içinde (393-436).

Aydın E., & Önder O.(2010). Sınava Hazırlık Biçiminin Farklı Sınav Türlerinde Ölçülen Matematik Sınav Başarı Düzeylerine Etkisi. *M.Ü. Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, 31, 5-24

Aydın, A., Sarıer, Y., & Uysal, Ş. (2012). Sosyoekonomik ve sosyokültürel değişkenler açısından PISA matematik sonuçlarının karşılaştırılması. *Eğitim ve Bilim*, 37,164-172.

Baki, A. (2018). *Matematiği öğretme bilgisi*(3. Baskı). Pegem Akademi.

Başol, G. (2014). *Ölçek geliştirme çalışmalarının içerik analizive metodolojik değerlendirmesi. Türkiye örneklemi öğrencilerin akademik başarıları üzerindeki etkisi* [Sözlü bildiri]. 6. Uluslararası Eğitim Araştırmaları Kongresi, Ankara.

Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması* (2.Baskı). Pegem Akademi.

Ben-Simon, A. & Cohen, Y. (2004). *International assessments:* Merits and pitfalls. Proceedings of the 30th Annual Conference of the International Association for Educational Assessment (IAEA).

Berberoğlu, G., Demirtaşlı, N., İş Güzel, C., Arıkan, S., & Özgen Tuncer, C. (2010). Okul Dışı Etmenlerin Öğrenci Başarısı ile İlişkisi [Effects of out-of-school factors on students' achievement]. Cito Eğitim: Kuram ve Uygulama, 1(7), 28-38.

Bible, L., Simkin, M. G. & Kuechler,L., (2008). *Using Multiple-choice Tests to Evaluate Students' Understanding of Accounting Accounting Education 17, An İnternational Jour*nal, 17(1), S55-S68. DOI: 10.1080/09639280802009249

Bilgeç, İ. (2016). *Açık uçlu soruların kullanıldığı matematik sınavlarının ölçme ve değerlendirme perspektifinden incelenmesi* (Master's thesis, Eğitim Bilimleri Enstitüsü).

Birgin, O. (2010). *4 - 5. Sınıf Matematik Öğretim Programında Öngörülen Ölçme ve Değerlendirme Yaklaşımlarının Öğretmenler Tarafından Uygulanabilirliği*" (Doktora Tezi), Karadeniz Teknik Üniversitesi.

Birgili, B., (2006). *Open Ended Questıons As An Alternatıve To Multıple Choıce: Dılemma In Turkısh Examınatıon System* (yayınlanmamış Yüksek Lisans Tezi). ODTÜ.

Bozkurt, B. Ü. (2016). T*ürkiye'de okuma eğitiminin karnesi: pisa ölçeğinden çıkarımlar.* Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi, 16 (4), 1673-1686. http://dx.doi.org/10.17240/aibuefd.2016.16.4-5000209673

Bozkurt, E., & Dülger, E. (2021). Matematik eğitiminde ebeveynlerin rolü. *Cumhuriyet International Journal of Education, 10*(4), 1507-1534. http://dx.doi.org/10.30703/cije.873357

Büyüköztürk, Ş. (2012). *Sosyal bilimler için veri analizi el kitabı*. Pegem Akademi.

Büyüköztürk, S., Kılıç Çakmak, E., Akgün, O. E., Karadeniz, S. & Demirel F. (2015). *Bilimsel araştırma yöntemleri*. Pegem Akademi.

Cambaz, G. (2000). *İlköğretim Okullarının Beşinci ve Sekizinci Sınıflarında Çoktan Seçmeli Test Sınavı ile Klasik Yazılı Sınavın Matematik Öğretimi ile İlişkisi*. (Yayımlanmamış Yüksek Lisans Tezi), Marmara Üniversitesi.

Can, A., (2014). *SPSS İle Bilimsel Araştırma Sürecinde Nicel Veri Analizi,* Pegem Akademi.

Chaoui, N. A. (2011). Finding relationships between multiple-choice math tests and their stem-equivalent constructed responses. The Claremont Graduate University.

Cohen, J. (1988). *The t test For Means. Statistical power analaysis fort he behavioural sciences*. Hillsdale, NJ, Earlbaum, 567.

Çepni, S., & Çil, E. (2009). A case study for reliability on portfolio evaluation. *Procedia-Social and Behavioral Sciences*, *1*(1), 963-968.

Çelen, F. K., Çelik A. & Seferoğlu, S. S., (2011, 2 – 4, Şubat)Türk *eğitim sistemi ve pisa sonuçları.* 2011Akademik Bilişim'11 - XIII. Akademik Bilişim Konferansı Bildirileri İnönü Üniversitesi, Malatya

Demir, E. (2010). Uuslararası öğrenci değerlendirme programı (pisa) bilişsel alan testlerinde yer alan soru tiplerine göre türkiye'de öğrenci başarıları (Yayımlanmamış Yüksek Lisans tezi). Hacettepe Üniversitesi.

Demirbilek, S. (2015). *Öğretmen adaylarının eğitimde ölçme ve değerlendirme dersindeki kavram yanılgılarının incelenmesi (Yüksek Lisans).* Hacettepe Üniversitesi.

Demirtaşlı, N. (2010, Nisan-Haziran). *Açık uçlu soru formatı ve öğrenci izleme sistemi (öis) akademik gelişimi izleme ve değerlendirme (AGİD) modülündeki kullanımı*. Cito Eğitim: Kuram ve uygulama, 21-30.

Eren B. (2015) Çoktan seçmeli ve karma test uygulamalarına ilişkin öğrenci başarıları ile öğrenci ve öğretmen görüşlerinin karşılaştırılması. Ankara Üniversitesi, Ankara.

Gülteki, S.(2011). *Çoktan seçmeli, açık uçlu ve karma testlerin psikometrik özelliklerinin madde tepki kuramına dayalı olarak değerlendirilmesi* (Doktora Tezi), Ankara Üniversitesi.

Gültekin S., & N. Ç. Demirtaşlı, (2012). *Çoktan seçmeli, açık uçlu ve karma testlerden sağlanan bilginin madde tepki kuramına dayalı olarak karşılaştırılması.* İlköğretim Online, 11(1), 251-263.

Gündoğdu, K., & Ozan, C. (2013). Bilgisayar destekli öğretim. B. Oral (Ed.),Öğrenme öğretme kuram ve yaklaşımları içinde (2. baskı, ss. 385-410). Pegem Akademi.

Green, S. B. & Salkind, N. J. (2005). Using SPSS for *windows and macintosh:analyzing and understanding data* (4th edition) New Jersey. Pearson.

Haladavna T. M. (1997) Writina test items to evaluate higher order thinking

Herman, J.L., Klein, D.C., Heath, T.M. & Wakai, S.T. (1994). A first look: Are claims for alternative assessment holding up? (CSE Technical Report No.391). University of California, Center for Research on Evaluation Standards and Student Testing.

İlhan, M., Öztürk, NB & Şahin, MG. (2020). The effect of the item's type and cognitive level on its difficulty index: The sample of TIMSS 2015. *Participatory Educational Research*, *7*(2), 47-59. https://doi.org/10.17275/per.20.19.7.2

İskenderoğlu, T., & Baki, A. (2011). *İlköğretim 8. sınıf matematik ders kitabındaki soruların PISA matematik yeterlik düzeylerine göre sınıflandırılması*. *Education and Science*, *36*(161).

Kalaycı, Ş. (2009). SPSS uygulamalı çok değişkenli istatistik teknikleri (4. baskı). Asil Yayın.

Karaçam, S., *Öğrencilerin kuvvet ve hareket konularındaki kavramsal anlamalarının ve soru çözümünde kullandıkları bilişsel ve üstbilişsel stratejilerin soru tipleri dikkate alınarak incelenmesi*. (Yüksek Lisans), Gazi Üniversitesi, Ankara,

Karasar, N. (2009). Bilimsel araştırma yöntemi (20. baskı). Ankara: Nobel Yayın Dağıtım.

Kastnera, M., & Stangla, B., "Multiple Choice and Constructed Response Tests: Do Test Format and Scoring Matter?" Procedia Social and Behavioral Sciences 12 (2011) 263–273. https://doi.org/10.1016/j.sbspro.2011.02.035

Koğar, E., & Koğar, H. (2019). *Farklı madde türlerine göre fen okuryazarlığının incelenmesi: PISA 2015 TÜRKİYE örneği. Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, *19*(2), 695-709.

Koyuncu, E., & Özkan, Y. Ö. (2019). *Geniş Ölçekli Sınavlarda Açık Uçlu Soruların Kullanılmasına İlişkin Öğrenci Görüşlerinin İncelenmesi*. *Elektronik Sosyal Bilimler Dergisi*, *18*(69), 177-200.

Küçüker, N. (2017). The influence of item format on Turkish eighth graders' performance on the TIMSS.

Lyons-Thomas, J., Sandılands, D. & Ercikan K., "Gender Differential Item Functioning in Mathematics in Four International Jurisdictions" Education and Science, Vol 39, No 172, 2014. https://doi.org/10.15390/EB.2014.2846

MEB, (2019). TIMSS 2019 Türkiye Ön Raporu, Ankara.

National Council of Teachers of Mathematics. (1995). Assessment standards for school mathematics. Reston, VA: Author.

Üney, S.S, (2023) "Aynı özelliği ölçmeye yönelikolarak hazırlşanan çoktan seçmeli ve karma testlerin psikometrik özelliklerinin karşılaştırılması" (Yayımlanmamış Yüksek Lisans tezi). Hacettepe Üniversitesi.

Norman, G. (1995). Multiple choice questions evaluation methods. in evaluation methods: A Resource Handbook, The Program for Educational Development, Mc Master University. 47-54.

Oflaz, G. (2010). Geometrik düşünme seviyeleri ve zekâ alanları arasındaki ilişki [Yayımlanmamış yüksek lisans tezi]. Sivas Cumhuriyet Üniversitesi.

Özçelik, D. A. (2011). Ölçme ve değerlendirme (4.Baskı). Ankara: Pegem Akademi.

Öksüz, Y., & Güven Demir, E. (2019). Açık uçlu ve çoktan seçmeli başarı testlerinin psikometrik özellikleri ve öğrenci performansı açısından karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, *34*(1), 259-282. https://doi.org/10.16986/HUJE.2018040550

Umay, A. (1996). Matematik eğitimi ve ölçülmesi . *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 12, 145-149.

Schuwirth, L. W., & Van Der Vleuten, C. P. (2004). Different written assessment methods: what can be said about their strengths and weaknesses?. *Medical education*, *38*(9), 974-979. https://doi.org/10.1111/j.1365-2929.2004.01916.x

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics (6th ed.). Boston: Pearson

Toker, Turker, "A Comparison of Latent Class Analysis and the Mixture Rasch Model: A Cross-Cultural Comparison of 8th Grade Mathematics Achievement in the Fourth International Mathematics and Science Study (TIMSS-2011)" (2016). Electronic Theses and Dissertations. 1172.

Turgut, M. F. ve Baykul, Y. (2012). Eğitimde ölçme ve değerlendirme (4. Baskı). Ankara: Pegem Akademi.

Van den Akker, J. (2013). Curricular development research as a specimen. educational design research. In T. Plomp, & N. Nieveen (Eds.), Educational design research. Part A: An introduction, (pp. 52-71). Enschede: SLO.

Wainer, H., & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a Marxist theory of test construction. Applied Measurement in Education, 6(2), 103-118. https://doi.org/10.1207/s15324818ame0602_1

Wolf, P., Evers, F., & Hill, A. (2006). Handbook for curriculum assessment. Ontario: University of Guelph.

Yıldırım, Ö., Çıkrıkçı-Demirtaşlı, N. & Akbaş, U. (2012). The opinions of mathematics teachers on homework and ın-class assessment: TIMSS 1999 and TIMSS 2007 Periods, Education and Science, 37(163), 126-142.