



Açık Uçlu ve Çoktan Seçmeli Başarı Testlerinin Psikometrik Özellikleri ve Öğrenci Performansı Açısından Karşılaştırılması

Comparison of Open Ended Questions and Multiple Choice Tests In Terms Of Psychometric Features and Student Performance

Yücel ÖKSÜZ*, Elif GÜVEN DEMİR**

• *Geliş Tarihi:* 30.11.2017 • *Kabul Tarihi:* 17.04.2018 • *Yayın Tarihi:* 23.06.2018

ÖZ: Bu araştırmanın amacı açık uçlu başarı testi ile aynı testten yola çıkarak hazırlanan çoktan seçmeli başarı testini psikometrik özellikleri ve öğrenci performansı açısından karşılaştırmaktır. Bu amaçla ilkokul 4. Sınıf Fen Bilimleri ve Sosyal Bilgiler programlarında yer alan iki ünite kapsamında geliştirilen açık uçlu maddelerden oluşan testler dönüştürülerek çoktan seçmeli versiyonları oluşturulmuştur. Her ders için açık uçlu test ile çoktan seçmeli test madde güçlük, madde ayırıcılık, güvenilirlik, öğrenci performansı ve bilişsel alan basamakları açısından karşılaştırılmıştır. Araştırmanın verileri 2016-2017 eğitim öğretim yılında Samsun ilinde öğrenim gören 5. Sınıf öğrencileri ile yürütülmüştür. Araştırma sonuçları aynı üniteye yönelik olarak geliştirilen farklı formatlardaki testler arasında madde güçlük, madde ayırıcılık, öğrenci performansı, cinsiyet açısından anlamlı farklılık olduğunu göstermektedir.

Anahtar sözcükler: Açık Uçlu Test, Çoktan Seçmeli Test, Psikometrik, Bilişsel basamak

ABSTRACT: This study aimed to compare open ended questions test and modified test based on multiple choice questions in terms of psychometric features, student performance. Multiple choice items tests were modified from open ended items tests which were developed towards two units in 4th grade Science and Social Sciences lessons. Open ended questions tests and multiple choice questions tests for each lessons were compared in terms of item difficulty, item discrimination, reliability, student performance and cognitive levels. Research data was obtained from 5th grade secondary school students who study in Samsun at 2016-2017 academic year. Research results show that there was significant differences between different tests formats according to item difficulty, item discrimination, students performance, gender.

Keywords: Open ended questions test, Multiple choice test, Psychometric, Cognitive domain

1. GİRİŞ

Bireyin daha refah ve mutlu bir yaşam sürebilmesi için gerekli olan becerilerin kazandırılması, mevcut bilgi birikimlerinin yeni nesillere aktarılması çabası eğitim kavramına işaret etmektedir. Eğitim süreci ilkel toplumlardan günümüze kadar sürekli değişen ve insanlık tarihi kadar eski bir geçmişe dayanan dinamik ve çok yönlü bir yapıya sahiptir. Bireylerde oluşması beklenen davranış değişiklikleri, bunların kazandırılması ve sınanması işi eğitimin çok yönlü yapısının göstergeleridir.

Eğitim çıktılarının sınanması, eğitim süreç ve basamaklarının işlerliğinin kontrolünde doğal ve yadsınılamaz bir unsur olarak öne çıkmaktadır. Sınama işi ilk basamağı ölçme, ikinci basamağı ise mevcut ölçümlerin değerlendirilmesi olan iki aşamadan oluşmaktadır. Ölçme, bir niteliğin gözlenip gözlem sonucunun sayılarla veya başka sembollerle gösterilmesidir (Turgut, 1983). Değerlendirme ise ölçme sonuçlarının bir ölçüte vurarak, ölçülen nitelik hakkında bir değer yargısına varma sürecidir (Turgut,1977). Bu noktada ölçme ve değerlendirme eğitim sürecinin verimliliğine ve niteliğine ilişkin dönüt mekanizmasını oluşturmaktadır. Ölçme ve değerlendirme sürecinin çıktıları eğitimin niteliği, öğrenci performansı ve gelinen nokta

* Doç. Dr., Ondokuz Mayıs Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Samsun-TÜRKİYE. e-posta: yoksuz@omu.edu.tr (ORCID: 0000-0002-9310-7506)

** Arş. Gör., Ondokuz Mayıs Üniversitesi, Eğitim Bilimleri Enstitüsü, Samsun-TÜRKİYE. e-posta: elifguven1@hotmail.com (ORCID: 0000-0001-6685-5341)

hakkında önemli bir veri sunmaktadır. Böylece eğitimin ve ilgili boyutlarının değerlendirilmesi, düzeltilmesi ve geliştirilmesi olanağını sağlayan dinamik bir sürecin çekici kuvvetini oluşturmaktadır.

Eğitim çıktılarının, öğrenci performansını daha geçerli ve güvenilir yöntemlerle sınanması çabası çeşitli test tekniklerinin geliştirilmesine olanak sağlamıştır. Eğitimde kullanılan ölçme araçları bilişsel, duyuşsal ve devinişsel alana göre farklılık gösterirken, eğitim aşamasında ülkemizde genel sınavlar da dâhil her derste öğrencilerin bilişsel alan davranışlarını tespit etmede sıklıkla çoktan seçmeli testler tercih edilmektedir (Karamustafaoğlu ve Tutar, 2016).

Çoktan seçmeli testler cevapların cevaplayıcılar tarafından verilmediği, doğru cevabın testi hazırlayan tarafından seçenekler arasında verilerek, cevaplayıcılardan doğru cevabın bulunmasının istendiği testlerdir (Güler, 2017). Çoktan seçmeli testlerin bir takım avantaj ve dezavantajları olduğu söylenebilir. Kalabalık gruplarda puanlama güvenilirliğini sağlamak çoktan seçmeli testlerle daha kolaydır (Klufa, 2015). Çok sayıda maddeye yer verilebilmesi konu alanındaki kritik kazanımları kapsamına dolayısıyla da kapsam geçerliğinin yüksek olmasına imkan sağlamaktadır (Güler, 2017). Kısmi puanlama içermediği için daha objektif bir puanlama imkanı sunmaktadır (Turgut ve Baykul, 2012). Bunun yanı sıra çabuk puanlanabiliyor oluşu sebebiyle çoktan seçmeli testler okullara, öğretmenlere ve öğrencilere teşhis içerikli sonuç verebilmektedir (Birgili, 2014). Bununla birlikte çoktan seçmeli testlerde tahminle, eleme yöntemi ile doğru cevabı bulma ihtimali vardır. Seçenek sayısı arttıkça doğru yanıt tahminle bulma olasılığı azalmakla birlikte (Yazıcıoğlu, 2007); çoktan seçmeli testlerde öğrencilerin şans eseri doğru cevabı bulma olasılığı vardır (Gültekin, 2014).

Çoktan seçmeli testler, daha çok olgusal bilgiyi ölçmeye yönelik olup iyi bir şekilde hazırlansa bile ifade yeteneğinin dışında kalan yargılama, yorumlama, değerlendirme ve çözümlenme gibi üst düzeydeki zihinsel becerileri tam anlamıyla ölçemez (Üstüner ve Şengül, 2004). Çoktan seçmeli testlerde öğrencilerin işaretledikleri seçenekleri seçme gerekçeleri ve bunun altında yatan nedenleri belirlemek mümkün değildir (Karataş, Köse ve Coştu, 2003). Bunun yanı sıra çoktan seçmeli testlerde öğrencilerin kavram yanılgılarını hata ve eksik bilgiden ayırt etmek kolay değildir (Demirbilek, 2015). Bush (2001) çoktan seçmeli testlerin öğrencilerin gerçek beceri düzeyini ölçemeyebileceğini, öğrencilerin tahmin yolu ile yüksek puan alabileceğini ifade etmektedir. Nitekim öğrencilerin çoktan seçmeli testlerin kolay olduğunu düşündükleri ve derin çaba sarf etmedikleri ifade edilmektedir (Dochy, Janssens ve Struyven 2005). Çoktan seçmeli testlerde öğrenciler bilişsel stratejiler geliştirmek yerine, test çözme stratejilerine yatkın hale gelebilirler (Wolf, 1993; akt, Birgili, 2014). Entwistle ve Entwistle (1991) de çoktan seçmeli testlerin öğrencileri yüzeysel öğrenme eğilimine ittiğini belirtmektedir. Öğrenci performansı yalnızca doğru yanıtı göre değil; kullanılan strateji ve gerekçelendirme becerisine göre değerlendirilebilir (Birgili, 2014).

Uluslararası Öğrenci Değerlendirme Programı (PISA) gibi uluslararası sınavlarda soruların bir kısmı çoktan seçmeli maddelerden oluşurken diğer kısmında ise açık uçlu maddelere yer verilmektedir (PISA, 2015). PISA projesi öğrencilerin zorunlu eğitim sonunda ne kadar öğrendiklerini değil, edindikleri bu bilgi ve becerileri içinde yaşadıkları topluma katılırken ne kadar kullandıklarını ölçmeyi amaçlamaktadır (OECD, 2007). Bu noktada çeşitli üstün yanları ile açık uçlu maddelerin ön plana çıktığı söylenebilir.

Açık uçlu testler öğrencinin yanıtı kendisinin yapılandırdıkları testlerdir (Geer, 1988). Açık uçlu testlerde öğrenci, konuya yaklaşım biçiminde, cevabında kullanacağı olgusal bilgilerin seçiminde ve onları örgütleyip bütünleştirmede, cevabında kapsanan noktaların her birine vereceği önem derecesini belirlemede özgürdür (Temizkan ve Sallabaş, 2011).

Açık uçlu maddelerden oluşan testlerde öğrencinin algılayış, gerekçelendirme ve bilgiyi kullanma becerisine odaklanılır (Badger ve Thomas, 1992). Cooney, Sanchez, Leatham ve Mewborn (2004) öğrencilerin farklı şekillerde öğrendiklerini aynı zamanda bilgi ve düşüncelerini de farklı yöntemlerle gösterdiklerini, açık uçlu maddelerden oluşan bir testin öğrencilere bu imkanı tanıdığını ifade etmektedir. Bunun yanı sıra Bahar, Nartgün, Durmuş ve Bıçak (2012) açık uçlu maddelerin üst düzey becerileri ölçme konusunda çoktan seçmeli testlerden daha avantajlı olduğunu belirtmektedir. Açık uçlu maddelerde öğrencilerden bilinen bir tek çözümü olmayan eksik bilgili bir gerçek yaşam problemini çözmeleri istenerek öğrencilerin eksik bilgiler hakkında da kabuller ve yorumlar yaparak yaratıcı düşünce ve katkılarını ortaya koymaları sağlanabilir (Akay, Soybaş ve Argün, 2006).

Çoktan seçmeli testteki şans faktörü ve doğru cevabı eleyerek bulma ihtimalinin açık uçlu maddeler için söz konusu olmadığı söylenebilir. Lee, Liu ve Linn (2011) açık uçlu maddelerden elde edilen yanıtların öğretim süreci ve sürecin etkisini belirlemede daha kullanışlı olduğunu belirtmektedir. Benzer olarak Badger ve Thomas (1992) açık uçlu maddelerle yapılan bir testin öğrenci başarısı ve gerekçelendirme becerisinin yanı sıra öğretim yönteminin kalitesini de ortaya koyabileceğini de belirtmektedir.

Açık uçlu maddelerin de bir takım dezavantajları bulunmaktadır. Açık uçlu maddelere verilen cevapların uzun sürmesi çok sayıda soru sormayı engelleyerek, kapsam geçerliğini olumsuz etkileyebilir. Açık uçlu maddelerin puanlanması aşamasında, maddelere yönelik cevapların yazılı olarak verilmesi zorunluluğu, aslında ölçülmek istenmeyen özelliklerin de cevaba (dolayısıyla puanlamaya) yansımaya sebep olabilir (Temizkan ve Sallabaş, 2011).

Çoktan seçmeli ve açık uçlu maddelerin birbirlerine göre üstün yanları olduğu söylenebilir. Çoktan seçmeli ve açık uçlu testlerde maddede yer alan soru okunduğunda ve sorulara yanıt verirken aktif üretken, pasif otomatik etkileşim gibi farklı bilişsel süreçler devreye girmektedir (Vasan, Cheryl, DeFouw, Holland ve Vasan, 2017).

Literatürde açık uçlu testlerin çoktan seçmeli testlerle karşılaştırılmasına ilişkin çalışmalarda okuduğunu anlama (Rauch ve Hartig, 2010; Ko, 2010; Temizkan ve Sallabaş, 2011), anlamayı etkileme, öğrenci görüşleri (Duran ve Tufan, 2017); öğrenci performansı (Vasan vd, 2017), geçerlik ve güvenilirlik (Tekindal, 2000; Morgil ve Yılmaz, 2001), müdahale öncesi ve sonrası (Wooten, Cool, Prather ve Tanner, 2014) değişkenleri açısından incelendiği görülmektedir. Bu çalışmada ise açık uçlu maddelere dayalı test ile çoktan seçmeli testin psikometrik özellikleri ve öğrenci performansı açısından karşılaştırılması amaçlanmıştır. Bunun yanı sıra karşılaştırma çalışmalarında aynı konuya yönelik farklı test formatında farklı ya da benzer maddelerin hazırlandığı (Beller ve Gafni, 2000; Vasan vd, 2017) görülmektedir. Literatürden farklı olarak çoktan seçmeli maddeler farklı maddeler yerine açık uçlu maddelerden dönüştürülerek oluşturulmuştur. Bunun yanı sıra farklı test türündeki öğrenci performansı, madde güçlük ve ayırıcılık indeksleri bilişsel alan basamakları açısından karşılaştırılmıştır. Böylece bir test türünden farklı bir test türüne dönüştürüldüğünde maddenin güçlük ve ayırıcılık indekslerini karşılaştırma imkanı ortaya çıkmıştır. Bunun yanı sıra literatüre farklı test formatlarında üst düzey düşünme becerilerini ölçmeye yönelik öğrenci performansı, cinsiyet ve madde analizi perspektifinden karşılaştırılmalı bir veri sunulmaktadır. Ek olarak araştırma farklı içeriklere sahip iki ders üzerinden yürütülerek test türü değişkeni farklı içeriklere sahip dersler açısından incelenmiştir. Böylece farklı formattaki testlerin ölçme yetkinliklerine ilişkin alan literatürüne ve eğitimcilerin ölçme, değerlendirme tercihlerine katkı sunulması planlanmaktadır.

1.1. Araştırmanın Amacı

Bu araştırmanın amacı Fen Bilimleri ve Sosyal Bilgiler derslerine yönelik olarak hazırlanmış açık uçlu maddelerden oluşan başarı testi ile aynı maddelerden yola çıkarak hazırlanmış çoktan seçmeli başarı testini psikometrik özellikleri ve öğrenci performansı açısından karşılaştırmaktır. Bu genel amaçtan yola çıkarak aşağıdaki alt problemlere yanıt aranmıştır. Fen bilimleri ve Sosyal Bilgiler derslerine yönelik;

- Açık uçlu test ile çoktan seçmeli test arasında ortalama madde güçlük indeksi açısından anlamlı bir farklılık var mıdır?
- Açık uçlu test ile çoktan seçmeli test arasında ortalama madde ayırt edicilik indeksi açısından anlamlı bir farklılık var mıdır?
- Açık uçlu test ile çoktan seçmeli testten elde edilen puanların güvenilirlik katsayısı arasında anlamlı bir farklılık var mıdır?
- Açık uçlu test ile çoktan seçmeli test puan ortalamaları arasında anlamlı bir farklılık var mıdır?
- Açık uçlu ve çoktan seçmeli testlerin bilişsel alan basamakları puan ortalamaları arasında öğrenci performansı açısından anlamlı bir farklılık var mıdır?
- Açık uçlu ve çoktan seçmeli testlerin bilişsel alan basamakları puan ortalamaları arasında anlamlı bir ilişki var mıdır?
- Açık uçlu test ile çoktan seçmeli test puan ortalamaları arasında cinsiyet açısından anlamlı bir farklılık var mıdır?
- Açık uçlu ve çoktan seçmeli testlerin bilişsel alan basamakları puan ortalamaları arasında cinsiyet değişkeni açısından anlamlı bir farklılık var mıdır?

2. YÖNTEM

2.1. Araştırma Modeli

Açık uçlu maddelerden oluşan başarı testi ile çoktan seçmeli maddelerden oluşan başarı testinin psikometrik özellikleri, öğrenci başarı düzeyi açısından karşılaştırılmasına yönelik bu çalışma betimsel tarama modelinde hazırlanmıştır.

2.2. Evren ve Örneklem

Araştırmanın ulaşılabilir evrenini 2016- 2017 eğitim öğretim yılında Samsun iline bağlı okullarda okuyan 5. Sınıf öğrencileri oluşturmaktadır. Ölçme araçlarının hazırlandığı üniteler 4. Sınıf düzeyinde son üniteler olması sebebiyle 4. Sınıf öğrencileri yerine bu üniteleri işlemiş olan 5. Sınıf öğrencileriyle yapılmıştır.

Araştırma örneklemini iki yöntemle belirlenmiştir. Okulun belirlenmesinde farklı sosyo ekonomik düzeyde öğrencilerin öğrenim gördüğü ve civar bölgelerden göç alan okulun seçilmesi yoluna gidilerek uygun örnekleme yöntemi uygulanmıştır. Testlerin uygulandığı sınıfların seçiminde ise basit seçkisiz örnekleme yöntemi tercih edilerek Atakum ilçesinde yer alan bir devlet okulunda öğrenim gören 52 kız, 50 erkek olmak üzere toplam 102 5. Sınıf öğrencisi örnekleme dahil edilmiştir.

2.3. Veri Toplama Araçları

Verilerin toplanmasında 4. Sınıf Fen Bilimleri ve Sosyal Bilgiler dersi kapsamında geliştirilen açık uçlu maddelerden oluşan başarı testi ve bu testte yer alan maddelerden oluşturulan çoktan seçmeli test kullanılmıştır. Fen Bilimleri dersine yönelik ölçümler Dünya ve Evren ünitesi kapsamında geliştirilen testlerle yapılmıştır. Sosyal Bilgiler dersine yönelik ölçümler ise Uzaktaki Arkadaşım ünitesi kapsamında geliştirilen testlerle yapılmıştır. Başka bir ifade ile çalışma kapsamında her ders için iki olmak üzere toplam 4 test kullanılmıştır.

Açık uçlu maddelerden oluşan testlerin kapsam geçerliğinin hesaplanmasında uzman görüşüne başvurulmuştur. Fen Bilimleri dersi açık uçlu test için Fen eğitimi alanında çalışan 1 Profesör, 1 Doçent, 6 Araştırma görevlisi (Dr.) olmak üzere 8 alan uzmanına görüşleri için başvurulmuştur. Fen bilimleri testinde yer alan 16 maddeden üçü için düzeltilmesi yönünde bir uzman görüşü bulunmaktadır. Kapsam geçerlik oranının belirlenmesinde (Lawsh, 1975)'nin

$$KGO = \frac{N_G - \frac{N}{2}}{\frac{N}{2}}$$

formülü kullanılmıştır. KGO değerlerinin Lawshe (1975) tarafından belirlenen 0.05 anlamlılık düzeyindeki minimum kritik değerleri için Wilson, Pan ve Schumsky (2012) tarafından revize edilen ve 8 uzman için minimum kritik değer olan 0.693 referans alınmıştır. Buna göre Fen Bilimleri dersine ilişkin açık uçlu test için KGO ortalamaları üzerinden hesaplanan Kapsam Geçerliği İndeksi 0.95 olarak hesaplanmıştır. Bu değer 8 uzman için belirtilen kapsam geçerlik ölçütü olan .693'den büyük olduğu için ($KGİ \geq KGO$) oluşturulan testin kapsam geçerliğinin istatistiksel olarak anlamlı olduğu tespit edilmiştir.

Sosyal Bilgiler dersine ilişkin açık uçlu test için Sosyal Bilgiler alanında çalışan 3 Profesör, 1 Doçent, 1 Yardımcı Doçent, 3 Araştırma görevlisi (Dr.) olmak üzere 8 alan uzmanına görüşleri için başvurulmuştur. Sosyal bilgiler testinde yer alan 27 maddeden dördü için düzeltilmesi yönünde uzman görüşü alınmıştır. Lawshe (1975) tekniği kullanılarak hesaplanan kapsam geçerlik oranlarının ortalaması olan kapsam geçerlik indeksi 0.96 olarak hesaplanmıştır. Kapsam geçerlik oranı 8 uzman için minimum kritik değer olan 0.693' ten büyük olduğu için oluşturulan testin kapsam geçerliği istatistiksel olarak anlamlı olduğuna kanaat getirilmiştir.

Açık uçlu test maddeleri yanıtları kısmi puanlı 0, 1, 2 şeklinde kategorilendirilerek puanlanmıştır. Açık uçlu teste ilişkin öğrenci yanıtlarının puanlanmasında her ders için ayrı hazırlanan analitik dereceli puanlama anahtarı kullanılmıştır.

Fen Bilimleri dersine yönelik puanlama anahtarı 16 bireysel puanlama anahtarından oluşmaktadır. Her puanlama anahtarı kendi içinde Tam Doğru, Kısmen Doğru ve Yanlış/Boş Yanıt şeklinde gruplandırılmıştır. Tam Doğru: 2, Kısmen Doğru:1, Yanlış/Boş Yanıt: 0 şeklinde puanlanmıştır. Öğrenci yanıtlarına verilebilecek ihtimal yanıtlar örneklendirilerek ifade edilmiştir. Dereceli puanlama anahtarının maddeleri karşılama düzeyi, yanıtların ifade ediliş biçimi ve maddelere ilişkin olası yanıtlar için Fen eğitimi alanında çalışan 1 Profesör, 1 Doçent, 3 Araştırma Görevlisi (Dr.) olmak üzere 5 uzmanın görüşüne başvurulmuştur. Lawshe (1975) tekniği kullanılarak uzman görüşleri doğrultusunda hesaplanan KGİ değerinin (0.95) Wilson, Pan ve Schumsky (2012) tarafından revize edilen minimum değerlere göre 5 uzman için belirlenmiş kapsam geçerlik minimum ölçütü olan 0.87'yi aştığı tespit edilmiştir. Buna göre Fen Bilimleri dersi analitik dereceli puanlama anahtarının kapsam geçerliğinin istatistiksel olarak anlamlı olduğuna kanaat getirilmiştir.

Fen Bilimleri dersi analitik dereceli puanlama anahtarının güvenilirliğinin sağlanması noktasında puanlayıcılar arası güvenilirlik çalışmasına yer verilmiştir. Seçkisiz yöntemle 102 öğrenci arasından rastgele seçilen 63 öğrenciye ait açık uçlu testler dereceli puanlama anahtarı kullanılarak 2 puanlayıcı tarafından ayrı ayrı puanlanmıştır. Puanlayıcılar arası uyumun göstergesi olarak Pearson korelasyon katsayısı ve Krippendorff'un Alfa katsayısı hesaplanmıştır. İki puanlayıcının puanları arasında ayrı ayrı maddeler açısından ve testin geneli açısından pozitif yönde yüksek düzeyde anlamlı ilişki olduğu tespit edilmiştir ($r=.959$, $p<.01$). Krippendorff'un Alfa katsayısının ise 0.959 düzeyinde olduğu ve bu değer puanlayıcılar arasında çok yüksek düzeyde uyuma işaret ettiği tespit edilmiştir (Krippendorff, 2007). Bu verilerden hareketle Fen Bilimleri dersi analitik dereceli puanlama anahtarının güvenilir bir ölçme aracı olduğuna kanaat getirilmiştir. Fen Bilimleri dersine yönelik açık uçlu testte 16 madde yer almakta olup, testten alınabilecek en düşük puan 0, en yüksek puan ise 32'dir.

Sosyal Bilgiler dersine yönelik puanlama anahtarı 27 bireysel puanlama anahtarından oluşmaktadır. Her puanlama anahtarı kendi içinde Tam Doğru, Kısmen Doğru ve Yanlış/Boş Yanıt şeklinde gruplandırılmıştır. Tam Doğru: 2, Kısmen Doğru:1, Yanlış/Boş Yanıt: 0 şeklinde

puanlanmıştır. Puanlama anahtarının kapsam geçerliğinin sağlanması amacıyla Sosyal Bilgiler alanında çalışan 3 Profesör, 1 Doçent ve 1 Araştırma Görevlisi (Dr.) olmak üzere 5 uzmanın görüşüne başvurulmuştur. Uzman görüşleri doğrultusunda hesaplanan KGİ değerinin (.89) kritik minimum değer olan .87'yi aşarak .05 düzeyinde anlamlı olduğu tespit edilmiştir.

Sosyal Bilgiler analitik dereceli puanlama anahtarının güvenilirliğinin sağlanmasında Fen Bilimleri analitik dereceli puanlama anahtarında olduğu gibi puanlayıcılar arası güvenilirlik çalışmasına yer verilmiştir. Seçkisiz yöntemle rastgele seçilen 63 öğrenciye ait açık uçlu testler 2 puanlayıcı tarafından ayrı ayrı puanlanmıştır. Puanlayıcılar arası uyumu işaret eden Pearson Korelasyon kat sayısı .948, Krippendorff Alfa kat sayısı .947 olarak hesaplanmıştır. Elde edilen verilerden hareketle Sosyal Bilgiler Dersi analitik dereceli puanlama anahtarının güvenilir bir ölçme aracı olduğu kanaatine varılmıştır. Sosyal Bilgiler dersine yönelik açık uçlu testte 27 madde yer almakta olup, testten alınabilecek en düşük puan 0, en yüksek puan ise 54'tür.

Hazırlanan açık uçlu testlerin çoktan seçmeli maddelere dönüştürülmesinde ise farklı formattaki maddelerin aynı yapıyı ölçüp ölçmediklerine ilişkin uzman görüşüne başvurulmuştur. 3 uzman tarafından Fen Bilimleri dersi için çoktan seçmeli testteki 2 maddenin kökünde düzeltme yapılması yönünde görüş bildirilmiştir. Sosyal Bilgiler dersi için açık uçlu maddelerle çoktan seçmeli testteki maddelerin aynı yapıyı ölçtüklerine yönelik görüş bildirilmiştir. Uzmanlardan alınan dönütlere göre Fen Bilimleri ve Sosyal Bilgiler derslerine ilişkin testlerde düzenlemeler yapılmıştır. Çoktan seçmeli testler açık uçlu testlerde yer alan maddelerden dönüştürülerek hazırlandığı için tekrar kapsam geçerliği çalışmasına yer verilmemiştir. Çoktan seçmeli testlerin seçenek sayısının belirlenmesinde sınıf düzeyi dikkate alınmıştır. İlköğretim basamağındaki çoktan seçmeli testlerde 3 ve ya 4 seçenekli madde türlerinin tercih edildiği belirtilmektedir (Yaman, 2016; Baştürk, 2014). Sınıf düzeyine uygun olmasının yanı sıra, şans başarısını azaltmak için çoktan seçmeli testlerde 4 seçenekli maddelere yer verilmiştir.

Fen Bilimleri dersi çoktan seçmeli testinde yer alan madde sayısı 16 iken; Sosyal Bilgiler dersinde 27'dir. Çoktan seçmeli maddeler için 1,0 puanlama yapılmıştır. Buna göre; Fen Bilimleri dersi çoktan seçmeli testinden alınabilecek en düşük puan 0, en yüksek puan ise 16'dır. Sosyal Bilgiler dersinde ise alınabilecek en düşük puan 0, en yüksek puan 27'dir.

2.4. Verilerin Toplanması

Çoktan seçmeli testlerde öğrencilerin seçeneklerden yola çıkarak yanıt bulma ihtimaline karşı ve kendi yanıtlarını yapılandırmalarını gerektirmesi sebebiyle çalışma grubuna öncelikle açık uçlu testler uygulanmıştır. İki uygulama arasında süre olmaması için çoktan seçmeli testler açık uçlu testlerin hemen ardından uygulanmıştır. Fen Bilimleri dersine ilişkin testlerin uygulanması yaklaşık 2 ders saati sürerken, Sosyal Bilgiler dersine yönelik testler ise yaklaşık 4 ders saatinde uygulanmıştır. Öğrencilerde oluşabilecek yorgunluk, bıkkınlık ihtimaline karşı, uygulamalar yalnızca ders saatleri içinde yapılmış; öğrencilerin tenefüs arası vermelerine izin verilmiştir. Tenefüs aralarında oluşabilecek kopya ihtimaline karşı Fen Bilimleri dersine ilişkin testler 4 ayrı formda düzenlenmiştir. Buna göre, ilk ders öğrencilere öncelikle 8 açık uçlu maddenin yer aldığı ilk form uygulanmıştır. Hemen ardından ise aynı maddelerin çoktan seçmeli versiyonlarının yer aldığı ikinci form uygulanmıştır. Sonraki derste ise kalan açık uçlu maddeler ve çoktan seçmeli versiyonlarından oluşan formlar öğrencilere uygulanmıştır. Böylece, testlerdeki madde sayısından kaynaklı uzun süren yanıtlama süresinin sınırlılığı tenefüslerle; tenefüs aralarının getirdiği kopya ihtimali sınırlılığı ise form sayısı artırılarak aşmaya çalışılmıştır. Sosyal Bilgiler dersine yönelik testler, Fen Bilimleri dersinde olduğu gibi form sayısı artırılarak ve yalnızca ders saatini alacak şekilde uygulanmıştır. Sınav psikolojisinin öğrencilerde oluşturabileceği bıkkınlık ve yorgunluk ihtimaline karşı her dersin uygulaması farklı bir günde yapılmıştır.

2.5. Verilerin Analizi

Testlere ilişkin madde güçlük ve madde ayırıcılık indekslerinin karşılaştırılmasında oran karşılaştırmalarında kullanılan z testinden, Cronbach Alpha ve KR-20 güvenirlik katsayılarının karşılaştırılmasında Alsawalmeh ve Feldt (1994) tarafından önerilen formülle ölçülebilen Feldt testinden faydalanılmıştır.

$$W = \frac{(1 - \alpha_1)}{(1 - \alpha_2)}$$

α_1 : 1. teste yönelik güvenirlik katsayısı

α_2 : 2. teste yönelik güvenirlik katsayısı

Alsawalmeh ve Feldt (1994) tarafından $(k - 1).(N - 1)$ değeri 1000'in üzerine çıktığı durumlarda bağımlı gruplar için bu formülün kullanılabileceği, elde edilen W değerinin, $F(N-1, N-1)$ serbestlik derecesi değeri ile karşılaştırılması gerektiği belirtilmiştir (k: tüm ölçme araçlarındaki madde sayısı). Açık uçlu ve çoktan seçmeli teste yönelik güvenirlik karşılaştırılmasında $(k-1).(N-1)$ değerinin 1000'in üzerinde olması üzerine bu formülün kullanılmasına karar verilmiştir.

Verilerin analizinde kullanılacak testlere karar verirken parametrik test varsayımlarından biri olan verilerin normal dağılım gösterme durumu test edilmiştir. Testlerin farklı kategorilerde puanlanıyor olması sebebiyle test puanları Z puanına dönüştürülerek standartlaştırılmıştır. Elde edilen z puanlarından hareketle uç değer tespiti yapılmıştır. Puanların normal dağılımdan önemli bir sapma göstermediğinin bir işareti olması sebebiyle çarpıklık ve basıklık değerlerinin +1 ve -1 aralığında olup olmadığı, kontrol edilmiştir (Büyüköztürk, Çokluk ve Köklü, 2010). Bunun yanı sıra Hair, Black, Babin, Anderson ve Tatham (2006) normal dağılım varsayımının testinde + 2.58 (.01 anlamlılık düzeyi için) ve +1.96 (.05 anlamlılık düzeyi için) kritik değerleri arasında yer alan z istatistiğinin referans alınabileceğini ifade etmektedir. Benzer olarak Büyüköztürk, Çokluk ve Köklü (2010), çarpıklık katsayısının standart hatasına bölünmesi ile elde edilen z istatistiği değerinin $\alpha=.05$ için 1.96'dan küçük çıkması durumunda dağılımın normalden aşırı sapma göstermediği şeklinde yorumlanabileceğini ifade etmektedir. Normal dağılımın test edilmesinde spesifik normallik testlerinin de kullanılabileceği ifade edilmektedir (Hair vd.,2006). Bu noktadan hareketle Kolmogorov Smirnov testine başvurulmuştur.

Tablo 1: Fen bilimleri ve sosyal bilgiler testlerine ilişkin çarpıklık değerleri ve Kolmogorov Smirnov testi sonuçları

Ders	Testin Türü	Kolmogorov-Smirnov			Çarpıklık		z	
		İstatistik	Sd	p	İstatistik	Std.Hata		
Fen Bilimleri	Açık Uçlu Test	Kız	.085	52	.200*	.248	.330	.751
		Erkek	.107	50	.200*	.161	.337	.477
	• Anlama Basamağı	Kız	.142	52	.011	-.625	.330	-1.89
		Erkek	.074	50	.200*	-.710	.337	-2.10
	• Uygulama Basamağı	Kız	.121	52	.055	.364	.330	1.10
		Erkek	.143	50	.012	.019	.337	.563
	• Analiz Basamağı	Kız	.079	52	.200*	.292	.330	.884
		Erkek	.071	50	.200*	.151	.337	.448
	Çoktan Seçmeli Test	Kız	.104	52	.200*	-.522	.330	-1.58
		Erkek	.118	50	.079*	.133	.337	.394
	• Anlama Basamağı	Kız	.254	52	.000	-.607	.330	-1.83
		Erkek	.213	50	.000	-.745	.337	-2.21
	• Uygulama Basamağı	Kız	.193	52	.000	-.552	.330	-1.67
		Erkek	.133	50	.026	-.556	.337	-1.64
	• Analiz Basamağı	Kız	.184	52	.000	.278	.330	.842
		Erkek	.150	50	.006	-.415	.337	-1.23

Sosyal Bilgiler	Açık Uçlu Test	Kız	.117	53	.069*	-.569	.327	-1.74
		Erkek	.059	49	.200*	-.147	.340	-0.432
	• Anlama Basamağı	Kız	.095	53	.200*	-.299	.327	-.914
		Erkek	.131	49	.036	-.315	.340	-.926
	• Uygulama Basamağı	Kız	.187	53	.000	-.439	.327	-1.34
		Erkek	.119	49	.078*	-.081	.340	-.238
	• Analiz Basamağı	Kız	.143	53	.009	-.736	.327	-2.25
		Erkek	.117	49	.091*	-.594	.340	-1.74
	Çoktan Seçmeli Test	Kız	.146	53	.007	-.520	.327	-1.59
		Erkek	.103	49	.200*	-.701	.340	-2.06
	• Anlama Basamağı	Kız	.223	53	.000	.661	.327	2.02
		Erkek	.149	49	.015	-.554	.340	-1.62
	• Uygulama Basamağı	Kız	.272	53	.000	.434	.327	1.32
		Erkek	.217	49	.000	.310	.340	.911
	• Analiz Basamağı	Kız	.257	53	.000	-.658	.327	-2.01
		Erkek	.144	49	.013	-.300	.340	-.882

*p>.05

Tablo 1 incelendiğinde Fen Bilimleri dersine yönelik açık uçlu ve çoktan seçmeli test puanlarının kız ve erkek öğrenciler arasındaki dağılımında Kolmogorov Smirnov testi null hipotezinin kabul edildiği görülmektedir ($p>.05$). Bununla birlikte açık uçlu testte analiz basamağı dışında ($p>.05$), iki test türünde bilişsel basamak testleri açısından null hipotezinin reddedildiği görülmektedir ($p>.05$). Sosyal Bilgiler dersinde ise açık uçlu test puanlarının dağılımı için null hipotezinin kabul edildiği ($p>.05$), çoktan seçmeli test için reddedildiği görülmektedir ($p<.05$). Çalışmada yer alan testler çarpıklık katsayısının standart hatasına bölünmesi ile elde edilen z istatistiği açısından incelendiğinde, bu değerlerin $\alpha=.05$ için 1.96 ve $\alpha=.01$ için 2.58'den küçük oldukları tespit edilmiştir. Bu durumun puanların normal dağılıma yaklaşık dağıldığı şeklinde yorumlanabileceği referansından hareketle, verilerin analizinde parametrik testlerin kullanılabilmesine kanaat getirilmiştir (Büyüköztürk vd., 2010).

Açık uçlu ve çoktan seçmeli testlerin karşılaştırılmasında bağımlı gruplar t testine başvurulmuştur. Açık uçlu ve çoktan seçmeli testlerin cinsiyet değişkenine göre incelenmesinde ise iki yönlü ANOVA testine başvurulmuştur. Kovaryans matrislerinin eşitliği ve bağımlı değişkenlerin varyanslarının homojenliği Box'ın M ve Levene'in F testleri incelenerek sınanmıştır. İki yönlü ANOVA testinin varsayımlarından olan küresellik varsayımı Mauchly testi ile test edilmiş; ancak sonucun anlamı çıkması üzerine Greenhouse Geisser F değeri referans alınmıştır. Testlerin bilişsel alan basamakları arasındaki ilişkinin incelenmesinde ise Pearson korelasyon testine başvurulmuştur.

Gruplar arasındaki farkın anlamlı çıkması durumunda etki büyüklüğü eta-kare (η^2) değeri ve etki büyüklük indekslerinden birisi olan Cohen d indeksi incelenmiştir. Elde edilen eta kare ve Cohen d indeksleri aşağıdaki tabloya göre yorumlanmıştır (Cohen, 1988).

Tablo 2: Grup karşılaştırmalarında etki büyüklüğü yorum rehberi

Etki Büyüklüğü	Cohen d	Kısmi Eta Kare	Pearson r
Küçük	.2	.01	.10-.29
Orta	.5	.06	.30-.49
Büyük	.8	.138	.50 ve üstü

Testlere ilişkin madde güçlük ve madde ayırıcılık indeksleri arasındaki farka ilişkin etki büyüklüğünün hesaplanmasında, oranlar arasındaki etki büyüklüğünü belirlemeyi sağlayan Cohen'in h istatistiği kullanılmıştır (Cohen, 1988). Elde edilen h değerlerinin yorumu d değerlerinin yorumu gibidir (Kılıç ve Kelecioğlu, 2016). Güvenirlik değerlerinin karşılaştırılmasına ilişkin etki büyüklüğü ise Cohen'in d istatistiği ile belirlenmiştir.

3. BULGULAR

Bu bölümde verilerin analizi sonucunda elde edilen bulgulara yer verilmiştir. Fen Bilimleri ve Sosyal Bilgiler olmak üzere bulgular iki başlık halinde belirtilmiştir. Her ders için ayrı olmak üzere bulgulara araştırmanın alt problemleriyle aynı sıra ile yer verilmiştir.

3.1. Fen Bilimleri Dersine İlişkin Bulgular

Araştırma kapsamında açık uçlu ve çoktan seçmeli testlerin madde puanlarına dayalı test istatistikleri ve madde parametreleri (p_j , r_{jx}) belirlenmiş ve bunlar arasındaki farkın anlamlılığı test edilmiştir.

Tablo 3: Fen dersi açık uçlu test ve çoktan seçmeli test madde güçlük (p_j) ve madde ayırıcılık (r_j) düzeylerinin karşılaştırılması

Madde No	P_j^1	P_j^2	z	Cohen h	r_{jx}^1	r_{jx}^2	z	Cohen h
1	0.30 ^u	0.94 ^u	-9.41*	1.43	0.28 ^u	0.21 ^u	1.16	0.16
2	0.58 ^u	0.46 ^u	1.71	0.26	0.35 ^u	1 ^u	-9.91*	1.88
3	0.52 ^s	0.45 ^s	1.00	0.14	0.41 ^s	1 ^s	-9.23*	1.76
4	0.79 ^a	0.88 ^a	-1.73	0.22	0.21 ^a	0.42 ^a	-3.22*	0.44
5	0.31 ^s	0.33 ^s	-0.30	0.06	0.32 ^s	1 ^s	-10.2*	1.96
6	0.44 ^s	0.74 ^s	-4.356*	0.64	0.42 ^s	0.92 ^s	-7.59*	1.12
7	0.69 ^a	0.52 ^a	2.48*	0.36	0.26 ^a	0.96 ^a	-10.24*	1.62
8	0.70 ^u	0.81 ^u	-1.82	0.28	0.57 ^u	0.67 ^u	-1.47	0.18
9	0.10 ^u	0.62 ^u	-7.73*	1.14	0.32 ^u	1 ^u	-10.25*	1.96
10	0.90 ^u	0.91 ^u	-0.24	0.06	0.21 ^u	0.32 ^u	-1.78	0.24
11	0.50 ^a	0.75 ^a	-3.68*	0.52	0.55 ^a	0.92 ^a	-5.98*	0.84
12	0.63 ^a	0.74 ^a	-1.69	0.26	0.71 ^a	0.92 ^a	-3.86*	0.50
13	0.10 ^u	0.68 ^u	-8.49*	1.28	0.17 ^u	0.96 ^u	-11.38*	1.8
14	0.65 ^u	0.53 ^u	1.74	0.26	0.46 ^u	1 ^u	-8.68*	1.68
15	0.43 ^s	0.69 ^s	-3.74*	0.54	0.53 ^s	1 ^s	-7.91*	1.54
16	0.87 ^s	0.92 ^s	-1.16	0.12	0.19 ^s	0.28 ^s	-1.51	0.20
Anlama	0.65	0.72	-1.87	0.14	0.43	0.80	-5.43*	0.60
Uygulama	0.46	0.63	-2.43*	0.36	0.33	0.73	-5.72*	0.80
Analiz	0.51	0.62	-1.58	0.22	0.37	0.84	-6.86*	0.92
Test Ortalama	0.53	0.68	-2.19*	0.32	0.37	0.79	-6.07*	0.86

* $p < .05$

¹: Açık Uçlu Test ²: Çoktan Seçmeli Test

^a: Anlama Basamağı, ^u: Uygulama Basamağı, ^s: Analiz Basamağı

Tablo 3 incelendiğinde madde güçlük indekslerinin açık uçlu test için 0.10-0.90, çoktan seçmeli test için 0.33-0.94 aralığında değiştiği görülmektedir. Güler (2017) madde güçlük indeksinin 0-1 aralığında yer aldığı .50'nin ise orta nokta olduğunu belirtmektedir. Quairgrain ve Arhin (2017) madde güçlük indeksinin .20 ve altı: çok zor, .40-.60: orta, .90 ve üstü: çok kolay şeklinde yorumlanabileceğini ifade etmektedir. Buna göre; açık uçlu testte 2 adet çok zor, 2 adet zor, 5 adet orta güçlükte, 6 adet kolay, 1 adet çok kolay madde bulunmaktadır. Bununla birlikte çoktan seçmeli testte 1 adet zor, 4 adet orta güçlükte, 8 adet kolay, 3 adet çok kolay madde bulunmaktadır.

Bilişsel alan basamakları açısından incelendiğinde açık uçlu ve çoktan seçmeli testlerin anlama basamağındaki ortalama güçlük düzeyleri arasında anlamlı bir farklılığın olmadığı, anlama basamağının her iki test türünde de kolay olduğu tespit edilmiştir. Bununla birlikte anlama basamağındaki 7 ve 11 no' lu maddelerin madde güçlük düzeylerinin test türü değişkeni açısından anlamlı bir farklılık gösterdiği tespit edilmiştir. 7. maddenin açık uçlu testte kolay, çoktan seçmeli testte ise orta güçlükte, 11. maddenin ise açık uçlu testte orta güçlükte, çoktan seçmeli testte kolay bir madde olduğu tespit edilmiştir.

Açık uçlu ve çoktan seçmeli testlerin uygulama basamağındaki ortalama güçlük düzeyleri arasında anlamlı bir farklılığın olduğu, uygulama basamağındaki maddelerin ortalama güçlük düzeyinin açık uçlu testte orta, çoktan seçmeli testte ise kolay olduğu tespit edilmiştir. Test türü

değişkeninin uygulama basamağındaki ortalama madde güçlüğü üzerinde küçük etki düzeyine sahip olduğu söylenebilir. Bununla birlikte uygulama basamağındaki 1, 9 ve 13 no'lu maddelerin madde güçlük düzeylerinin test türüne göre anlamlı bir farklılık gösterdiği tespit edilmiştir. 1 no'lu maddenin açık uçlu testte zor, çoktan seçmeli testte çok kolay; 9 no'lu maddenin açık uçlu testte çok zor, çoktan seçmeli testte kolay; 13 no'lu maddenin açık uçlu testte çok zor, çoktan seçmeli testte kolay madde olduğu tespit edilmiştir. Etki büyüklükleri açısından incelendiğinde ise test türünün uygulama basamağındaki söz konusu maddelerin madde güçlük düzeyi üzerinde büyük etkiye sahip olduğu söylenebilir.

Açık uçlu ve çoktan seçmeli testlerin analiz basamağındaki ortalama güçlük düzeyleri arasında anlamlı bir farklılığın olmadığı görülmektedir. Bununla birlikte 6 ve 15 no'lu maddelerin madde güçlük düzeylerinin test türüne göre anlamlı farklılık gösterdiği tespit edilmiştir. 6 ve 15 no'lu maddelerin açık uçlu testte orta, çoktan seçmeli testte kolay olduğu tespit edilmiştir. Etki büyüklükleri açısından incelendiğinde ise test türünün analiz basamağındaki 6 ve 15 no'lu maddelerin madde güçlük düzeyi üzerinde orta büyüklükte etkiye sahip olduğu söylenebilir.

Açık uçlu testin ortalama güçlük düzeyi incelendiğinde testin orta güçlükte bir test olduğu görülmektedir. Çoktan seçmeli testin güçlük ortalaması ise testin kolay bir test olduğunu göstermektedir. Testlerin ortalama güçlük düzeyleri karşılaştırıldığında ise açık uçlu test ile çoktan seçmeli testin ortalama güçlük düzeylerinin anlamlı farklılık gösterdiği, çoktan seçmeli testin açık uçlu teste göre daha kolay bir test olduğu tespit edilmiştir. Testlerin güçlük düzeyleri arasındaki farkın 0.32 standart sapma kadar olduğu ve test türü değişkeninin test güçlüğü üzerinde küçük düzeyde etkiye sahip olduğu tespit edilmiştir.

Tablo 3' e göre testlerin madde ayırıcılık indekslerinin açık uçlu test için 0.17- 0.71, çoktan seçmeli test için 0.21-1 aralığında değiştiği görülmektedir. Quagrain ve Arhin (2017)' nin aktardığına göre madde ayırıcılık indeksi .19 ve altı: zayıf, .20-.29: kabul edilebilir, .30-.39: iyi .40 ve üstü : çok iyi düzeyde ayırt ediciliğe işaret etmektedir (Ebel, 1979). Buna göre; açık uçlu testte 2 adet çok zayıf, 4 adet kabul edilebilir, 3 adet iyi, 7 adet çok iyi düzeyde ayırıcılık düzeyine sahip maddenin yer aldığı söylenebilir. Çoktan seçmeli testte 2 adet kabul edilebilir, 1 adet iyi, 13 adet çok iyi düzeyde ayırt edici maddenin yer aldığı tespit edilmiştir.

Bilişsel alan basamakları açısından incelendiğinde açık uçlu ve çoktan seçmeli testlerin anlama basamağındaki ortalama ayırıcılık düzeyleri arasında anlamlı farklılığın olduğu görülmektedir. Her iki test türünde anlama basamağındaki ayırıcılık düzeyi ortalaması çok iyi düzeyde olmakla birlikte çoktan seçmeli testin bu basamakta ayırıcılık düzeyinin daha yüksek olduğu söylenebilir. Test türü değişkeninin anlama basamağındaki ortalama ayırıcılık düzeyi üzerinde orta düzeyde etkiye sahip olduğu tespit edilmiştir. Anlama basamağındaki maddeler ayrı ayrı incelendiğinde 4, 7,11 ve 12 no'lu maddelerin ayırıcılık düzeylerinin test türüne göre anlamlı farklılık gösterdiği, söz konusu maddelere ilişkin ayırıcılık düzeyinin çoktan seçmeli testte daha yüksek olduğu tespit edilmiştir. 4 no'lu maddenin açık uçlu testte zayıf, çoktan seçmeli testte çok iyi düzeyde; 7 no'lu maddenin açık uçlu testte zayıf, çoktan seçmeli testte çok iyi; 11 ve 12 no'lu maddelerin açık uçlu ve çoktan seçmeli testte çok iyi düzeyde ayırıcılık düzeyine sahip olduğu söylenebilir.

Açık uçlu ve çoktan seçmeli testlerin uygulama basamağındaki ortalama ayırıcılık düzeyleri arasında anlamlı farklılığın olduğu, açık uçlu testin bu basamaktaki ayırıcılık düzeyinin iyi, çoktan seçmeli testin ise çok iyi düzeyde olduğu tespit edilmiştir. Test türü değişkeninin uygulama basamağındaki ortalama ayırıcılık düzeyi üzerinde büyük bir etkisi olduğu tespit edilmiştir. Uygulama basamağında yer alan 2, 9, 13 ve 14 no'lu maddelerin farklı test türündeki madde ayırıcılık düzeyleri arasında anlamlı farklılık olduğu; söz konusu maddelere ilişkin madde ayırıcılık düzeylerinin çoktan seçmeli test türünde daha yüksek olduğu tespit edilmiştir.

Açık uçlu ve çoktan seçmeli testlerin analiz basamağındaki ortalama madde ayırıcılık düzeyleri arasında anlamlı farklılığın olduğu, test türü değişkeninin bu farklılık üzerinde büyük bir etkisinin olduğu tespit edilmiştir. Açık uçlu testin bu basamaktaki ayırıcılık düzeyinin orta, çoktan seçmeli testin ise iyi düzeyde olduğu tespit edilmiştir. 3, 5, 6 ve 15 no'lu maddelerin madde ayırıcılık düzeylerinin çoktan seçmeli testte açık uçlu teste göre daha yüksek olduğu tespit edilmiştir.

Testlerin geneli açısından incelendiğinde açık uçlu testin ortalama ayırıcılık düzeyinin iyi, çoktan seçmeli testin ortalama ayırıcılık düzeyinin ise çok iyi düzeyde olduğu söylenebilir. Testlerin ortalama ayırıcılık düzeyleri karşılaştırıldığında ise açık uçlu test ile çoktan seçmeli testin ortalama ayırıcılık düzeylerinin anlamlı farklılık gösterdiği, çoktan seçmeli testin açık uçlu teste göre daha ayırt edici bir test olduğu tespit edilmiştir. Testlerin ayırıcılık düzeyleri arasındaki farkın 0.86 standart sapma kadar olduğu ve test türü değişkeninin test ayırt ediciliği üzerinde büyük düzeyde etkiye sahip olduğu tespit edilmiştir.

Tablo 4 : Fen dersi açık uçlu test ve çoktan seçmeli test güvenilirlik katsayılarının karşılaştırılması

Testler	N	k	Cronbach Alpha /KR 20	W	Cohen d
Açık Uçlu Test	102	16	.661	0.954	0.137
Çoktan Seçmeli Test	102	16	.645		

p<.05

Tablo 4 incelendiğinde testlere ilişkin güvenilirlik katsayılarının 0.60'ın üstünde olduğu görülmektedir. Flynn, Schroeder ve Sakakibara (1994) 0.60 ve üstü değerlerin bir ölçme aracının güvenilirliği açısından yeterli olduğunu belirtmektedir (akt; Çelik; 2015). Bunun yanı sıra Ringim, Razallı ve Hasnan (2012)'in aktardığına göre Cronbach Alpha değeri 0.60' a kadar inebileceği ifade edilmektedir (Hair, Black, Babin ve Anderson; 2010). Buna göre testlere ilişkin güvenilirlik katsayılarının kabul edilebilir seviyede olduğu yorumu yapılabilir. Bunun yanı sıra Tablo 5'e göre açık uçlu test ile çoktan seçmeli testin güvenilirlik katsayıları arasında anlamlı bir fark bulunmamıştır ($W < F(101,101) = 1.38$). Buna göre test türü değişkeninin güvenilirlik düzeyi üzerinde etkili bir değişken olmadığı söylenebilir.

Tablo 5: Fen Bilimleri dersi açık uçlu test ve çoktan seçmeli teste ilişkin bağımlı gruplar t testi sonuçları

Testler	N	Ortalama	Std. Sapma	sd	t	p
Açık Uçlu Test	102	43.6	64	101	-2.934	.004*
Çoktan Seçmeli Test	102	58.5	61.1			

p<.05

Tablo 5 incelendiğinde öğrencilerin açık uçlu testten aldıkları puan ile çoktan seçmeli test puanları arasında anlamlı bir fark olduğu görülmektedir ($t=-2.934$, $p<.05$). Öğrencilerin çoktan seçmeli puan ortalamalarının açık uçlu test puan ortalamalarından daha yüksek olduğu tespit edilmiştir. Bu bulgu öğrencilerin çoktan seçmeli testte daha başarılı oldukları şeklinde yorumlanabilir. Öğrencilerin farklı test türündeki puan farklarına ilişkin etki büyüklüğü Cohen d

bağımlı gruplar için etki büyüklüğü formülü ($d = \frac{t}{\sqrt{N}}$) ile hesaplanmıştır. Yapılan hesaplama sonucunda Cohen d etki büyüklüğü indeksinin 0.29 olduğu ve bu değer küçük etki büyüklüğüne işaret ettiği tespit edilmiştir (Büyüköztürk, 2014). Bunun yanı sıra etki büyüklüğüne ilişkin Cohen d değeri Lenhard ve Lenhard, (2016) dönüşüm tablosuyla eta kare değerine dönüştürülmüştür ($\eta^2=0.02$). Etki büyüklüğüne ilişkin indeksler, öğrencilerin Fen Bilimleri dersi puanları arasındaki farkın yaklaşık 0.3 standart sapma kadar olduğunu, test türü değişkeninin küçük etki büyüklüğüne sahip olduğunu, öğrenci puanlarına ait varyansın %2'sinin test türü değişkeninden kaynaklandığını göstermektedir.

Tablo 6: Fen Bilimleri dersi açık uçlu test ve çoktan seçmeli test ölçümlerinin anlama, uygulama ve analiz basamağındaki puanlarının karşılaştırılmasına ilişkin bağımlı gruplar t testi sonuçları

	Testler	N	Ortalama	Std. Sapma	sd	t	p
Anlama Düzeyi	Açık Uçlu Test	102	49.1	24.4			
	Çoktan Seçmeli Test	102	52.7	22.4	101	-1.464	.146
Uygulama Düzeyi	Açık Uçlu Test	102	46.6	32.5	101	-2.495	.014*
	Çoktan Seçmeli Test	102	54.2	30.2			
Analiz Düzeyi	Açık Uçlu Test	102	47.8	25.8	101	-1.361	.177
	Çoktan Seçmeli Test	102	51.6	26.6			

p<.05

Tablo 6' ya göre öğrencilerin açık uçlu test anlama, uygulama ve analiz basamağı puanları ile çoktan seçmeli test anlama, uygulama ve analiz düzeyi puanları arasında anlamlı farklılık olmadığı tespit edilmiştir (t=-1.464, -2.495,-1.361; p>.01).

Tablo 7: Fen Bilimleri dersi açık uçlu test ve çoktan seçmeli test ölçümlerinin anlama, uygulama ve analiz basamağındaki puanları arasındaki ilişkinin incelenmesine yönelik Pearson korelasyon testi sonuçları

Testler	r	p	r ²
Anlama Basamağı (Açık Uçlu * Çoktan Seçmeli)	.454**	.000	.20
Uygulama Basamağı (Açık Uçlu * Çoktan Seçmeli)	.528**	.000	.27
Analiz Basamağı (Açık Uçlu * Çoktan Seçmeli)	.430**	.000	.18

**p<.01

Tablo 7 incelendiğinde açık uçlu ve çoktan seçmeli testlerin anlama basamağındaki puanları arasında pozitif yönde orta düzeyde bir ilişki olduğu görülmektedir (r=.454, p<.01, r²=.20). Uygulama basamağındaki puanlar arasında ise pozitif yönde orta düzeyde bir ilişki olduğu tespit edilmiştir (r=.528, p<.01, r²=.27). Analiz basamağında ise pozitif yönde orta düzeyde bir ilişki olduğu sonucuna ulaşılmıştır (r=.430, p<.01, r²=.18).

Tablo 8: Fen Bilimleri dersi açık uçlu ve çoktan seçmeli test puanlarının cinsiyete göre karşılaştırılmasına ilişkin iki yönlü ANOVA testi sonuçları

Varyansın Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p
Test Türü (Açık Uçlu-Çoktan Seçmeli)	11506.517	1	11506.517	8.803	.004
Test Türü* Cinsiyet	2115.512	1	2115.512	1.619	.206

p<.05

Tablo 8 incelendiğinde öğrencilerin açık uçlu ve çoktan seçmeli test puanlarının cinsiyet değişkeni açısından anlamlı farklılık göstermediği anlaşılmaktadır (F(1,2115.512)=1.619, p>.05).

Tablo 9: Fen Bilimleri dersi anlama, uygulama ve analiz düzeylerine ilişkin açık uçlu ve çoktan seçmeli test puanlarının cinsiyete göre karşılaştırılmasına ilişkin iki yönlü ANOVA testi sonuçları

	Varyansın Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p
Anlama Düzeyi	Test Türü (Açık Uçlu-Çoktan Seçmeli)	655.501	1	655.501	2.145	.146
	Cinsiyet	3081.807	1	3081.807	3.952	.051
	Test Türü* Cinsiyet	38.343	1	38.343	.125	.724
Uygulama	Test Türü (Açık Uçlu-	2899.343	1	2899.343	6.150	.015*

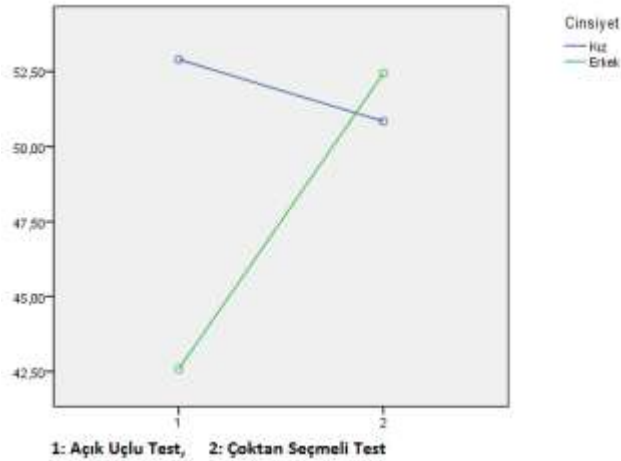
Düzeyi	Çoktan Seçmeli)					
	Cinsiyet	243.350	1	243.350	.160	.690
	Test Türü* Cinsiyet	7.529	1	7.529	.016	.900
	Test Türü (Açık Uçlu-Çoktan Seçmeli)	773.959	1	773.959	2.041	.156
Analiz Düzeyi	Cinsiyet	969.936	1	969.936	.983	.324
	Test Türü* Cinsiyet	1810.199	1	1810.199	4.774	.031*

* p<.05

Tablo 9 incelendiğinde anlama düzeyindeki test türü ana etkisinin ($F(1,655.501)=2.145$, $p>.05$), cinsiyet değişkeni ana etkisinin ($F(1,3081.807)=3.952$, $p>.05$) ve test türü* cinsiyet etkileşim etkisinin ($F(1,38.343)=.125$, $p>.05$) anlamlı olmadığı tespit edilmiştir. Buna göre kız ve erkek öğrencilerin anlama düzeyindeki açık uçlu ve çoktan seçmeli test puanlarının anlamlı farklılık göstermediği söylenebilir.

Uygulama düzeyinde ise test türü ana etkisinin anlamlı olduğu ($F(1, 2899.343)=6.150$, $p<.05$); ancak cinsiyet ana etkisinin ($F(1,243.350)=.160$, $p>.05$) ve test türü * cinsiyet etkileşim etkisinin ($F(1,7.529)=.016$, $p>.05$) anlamlı olmadığı tespit edilmiştir. Buna göre kız ve erkek öğrencilerin uygulama düzeyindeki açık uçlu ve çoktan seçmeli test puanları arasında anlamlı farklılık olmadığı söylenebilir.

Analiz düzeyi incelendiğinde test türü ana etkisinin ($F(1,773.959)=2.041$, $p>.05$) ve cinsiyet ana etkisinin ($F(1,969.936)=.983$, $p>.05$) anlamlı olmadığı; test türü*cinsiyet etkileşim etkisinin ise anlamlı olduğu tespit edilmiştir ($F(1,1810.199)=4.774$, $p<.05$, $\eta^2=.04$). Buna göre kız ve erkek öğrencilerin analiz düzeyindeki açık uçlu ve çoktan seçmeli test puanlarının farklılık gösterdiği; cinsiyet değişkeninin test türüne bağlı öğrenci performansı üzerinde küçük bir etkiye sahip olduğu söylenebilir.



Şekil 1: Kız ve erkek öğrencilerin Fen Bilimleri dersi analiz düzeyi açık uçlu ve çoktan seçmeli test ortalamaları

Şekil 1 incelendiğinde kız öğrencilerin analiz düzeyindeki açık uçlu test puan ortalamasının ($\bar{x}:52.5$) çoktan seçmeli test ortalamasından ($\bar{x}:50.8$) daha yüksek olduğu görülmektedir. Bununla birlikte erkek öğrencilerin analiz düzeyindeki çoktan seçmeli test puan ortalamasının ($\bar{x}:52.4$) açık uçlu test puan ortalamasından ($\bar{x}:42.5$) daha yüksek olduğu tespit edilmiştir. Açık uçlu test ile çoktan seçmeli test arasındaki ardışıklık durumu göz önüne alındığında kız öğrencilerin Fen Bilimleri dersi puanlarının düşüşte, erkek öğrencilerin puanlarının ise yükselişte olduğu görülmektedir. Kız öğrencilerin analiz düzeyinde açık uçlu testte çoktan seçmeli teste göre daha başarılı oldukları, erkek öğrencilerin ise analiz düzeyinde çoktan seçmeli testte açık uçlu teste göre daha başarılı oldukları söylenebilir.

3.2. Sosyal Bilgiler Dersine İlişkin Bulgular

Tablo 10: Sosyal bilgiler dersi açık uçlu test ve çoktan seçmeli test madde güçlük (p_j) ve madde ayırtıcılık (r_j) düzeylerinin karşılaştırılması

Madde No	P_j^1	P_j^2	z	Cohen h	r_{jx}^1	r_{jx}^2	z	Cohen h
1	0.62 ^a	0.95 ^a	-5.73*	0.88	0.42 ^a	0.21 ^a	3.22*	0.44
2	0.61 ^a	0.75 ^a	-2.14*	0.30	0.53 ^a	0.82 ^a	-4.42*	0.64
3	0.91 ^a	0.96 ^a	-1.44	0.16	0.16 ^a	0.14 ^a	0.4	0.06
4	0.77 ^a	0.97 ^a	-4.24*	0.62	0.51 ^a	0.10 ^a	6.35*	0.94
5	0.83 ^a	0.92 ^a	-1.94	0.22	0.30 ^a	0.28 ^a	0.31	0.04
6	0.61 ^a	0.90 ^a	-4.81*	0.66	0.71 ^a	0.35 ^a	5.15*	0.74
7	0.58 ^u	0.78 ^u	-3.06*	0.44	0.58 ^u	0.78 ^u	-3.06*	0.44
8	0.91 ^a	0.91 ^a	0	0	0.07 ^a	0.32 ^a	-4.50*	0.66
9	0.62 ^u	0.89 ^u	-4.48*	0.66	0.60 ^u	0.39 ^u	2.99*	0.42
10	0.67 ^a	0.97 ^a	-5.57	0.86	0.35 ^a	0.10 ^a	4.27*	0.62
11	0.84 ^u	0.92 ^u	-1.75	0.22	0.05 ^u	0.28 ^u	-4.42*	0.66
12	0.50 ^s	0.91 ^s	-6.42*	0.94	0.57 ^s	0.32 ^s	3.59*	0.52
13	0.59 ^s	0.85 ^s	-4.13*	0.62	0.23 ^s	0.53 ^s	-4.41*	0.62
14	0.79 ^a	0.57 ^a	3.36*	0.46	0.23 ^a	0.96 ^a	-10.6*	1.68
15	0.54 ^a	0.86 ^a	-4.98*	0.70	0.46 ^a	0.50 ^a	-0.57	0.10
16	0.86 ^s	0.93 ^s	-1.63	0.24	0.14 ^s	0.25 ^s	-1.98*	0.28
17	0.66 ^s	0.83 ^s	-2.78*	0.40	0.42 ^s	0.60 ^s	-2.57*	0.38
18	0.86 ^s	0.81 ^s	0.96	0.10	0.25 ^s	0.67 ^s	-6.01*	0.84
19	0.88 ^s	0.75 ^s	2.39*	0.30	0.23 ^s	0.89 ^s	-9.49*	1.46
20	0.46 ^a	0.74 ^a	-4.08*	0.62	0.46 ^a	0.92 ^a	-7.10*	1.04
21	0.87 ^a	0.87 ^a	0	0	0.16 ^a	0.46 ^a	-4.63*	0.64
22	0.76 ^u	0.93 ^u	-3.35*	0.46	0.46 ^u	0.25 ^u	3.13*	0.42
23	0.80 ^u	0.90 ^u	-2*	0.24	0.23 ^u	0.35 ^u	-1.88	0.98
24	0.54 ^u	0.81 ^u	-4.11*	0.60	0.41 ^u	0.67 ^u	-3.72*	0.50
25	0.53 ^u	0.84 ^u	-4.76*	0.68	0.66 ^u	0.57 ^u	1.32	0.18
26	0.93 ^u	0.68 ^u	4.50*	0.66	0.08 ^u	0.96 ^u	-12.5*	2
27	0.52 ^s	0.52 ^s	0	0	0.50 ^s	0.96 ^s	-7.33*	1.1
Anlama	0.71	0.86	-2.60*	0.35	0.36	0.43	-1.02	0.12
Uygulama	0.72	0.84	-2.06*	0.29	0.38	0.51	-1.86	0.26
Analiz	0.67	0.80	-2.10*	0.56	0.34	0.61	-3.86*	0.56
Test	0.71	0.84	-2.22*	0.28	0.36	0.50	2.01*	0.30
Ortalama								

* $p < .05$, ¹: Açık Uçlu Test, ²: Çoktan Seçmeli Test

^a: Anlama, ^u: Uygulama, ^s: Analiz

Tablo 10 incelendiğinde madde güçlük indekslerinin açık uçlu test için 0.46-0.93, çoktan seçmeli test için 0.52-0.97 aralığında değiştiği görülmektedir. Buna göre; açık uçlu testte 12 adet orta güçlükte, 5 adet kolay, 10 adet çok kolay madde yer almaktadır. Açık uçlu testin ortalama güçlük düzeyi incelendiğinde testin kolay bir test olduğu görülmektedir. Bununla birlikte çoktan seçmeli testte 2 adet orta güçlükte, 5 adet kolay, 20 adet çok kolay madde yer almaktadır. Çoktan seçmeli testin güçlük ortalaması ise testin çok kolay bir test olduğunu göstermektedir. Testlerin ortalama güçlük düzeyleri karşılaştırıldığında ise açık uçlu test ile çoktan seçmeli testin ortalama güçlük düzeylerinin anlamlı farklılık gösterdiği, çoktan seçmeli testin açık uçlu teste göre daha kolay bir test olduğu tespit edilmiştir. Testlerin güçlük düzeyleri arasındaki farkın 0.28 standart sapma kadar olduğu ve test türü değişkeninin test güçlüğü üzerinde küçük düzeyde etkiye sahip olduğu tespit edilmiştir.

Bilişsel alan basamakları açısından incelendiğinde anlama basamağındaki ortalama güçlük düzeyinin test türüne göre anlamlı farklılık gösterdiği, çoktan seçmeli testin açık uçlu teste göre daha kolay olduğu tespit edilmiştir. Test türü değişkeninin anlama basamağı ortalama güçlük düzeyi üzerinde küçük düzeyde etkisinin olduğu tespit edilmiştir. 1, 2, 4, 6, 15 ve 20 no'lu maddelerin istatistikî açıdan anlamlı bir şekilde çoktan seçmeli testte, 14 no'lu maddenin ise açık uçlu testte daha kolay olduğu tespit edilmiştir. Uygulama basamağındaki ortalama

güçlük düzeyinin test türü değişkenine göre anlamlı farklılık gösterdiği, çoktan seçmeli testin açık uçlu teste göre daha kolay olduğu tespit edilmiştir. Test türü değişkeninin uygulama basamağı ortalama güçlük düzeyi üzerinde küçük düzeyde etkisinin olduğu tespit edilmiştir. 7, 9, 22, 23, 24 ve 25 no'lu maddelerin anlamlı bir şekilde çoktan seçmeli testte, 26 no'lu maddenin ise açık uçlu testte daha kolay olduğu tespit edilmiştir. Analiz basamağındaki ortalama güçlük düzeyinin test türü değişkenine göre anlamlı farklılık gösterdiği, çoktan seçmeli testin açık uçlu teste göre daha kolay olduğu tespit edilmiştir. Test türü değişkeninin analiz basamağı ortalama güçlük düzeyi üzerinde orta düzeyde etkisinin olduğu tespit edilmiştir. 12, 13 ve 17 no'lu maddelerin anlamlı bir şekilde çoktan seçmeli testte, 19 no'lu maddenin ise açık uçlu testte daha kolay olduğu tespit edilmiştir.

Tablo 10' a göre testlerin madde ayırıcılık indekslerinin açık uçlu test için 0.05- 0.71, çoktan seçmeli test için 0.10-0.96 aralığında değiştiği görülmektedir. Buna göre; açık uçlu testte 6 adet çok zayıf, 5 adet düzeltilmeli, 2 adet iyi ve 14 adet çok iyi düzeyde ayırt edici maddenin yer aldığı görülmektedir. Testin ortalama ayırıcılık düzeyinin ise iyi düzeyde olduğu tespit edilmiştir. Çoktan seçmeli testte ise 3 adet çok zayıf, 5 adet kabul edilebilir, 5 adet iyi, 14 adet çok iyi düzeyde ayırt edici maddenin yer aldığı tespit edilmiştir. Testin ortalama ayırıcılık düzeyinin ise çok iyi düzeyde olduğu söylenebilir. Testlerin ortalama ayırıcılık düzeyleri karşılaştırıldığında ise açık uçlu test ile çoktan seçmeli testin ortalama ayırıcılık düzeylerinin anlamlı farklılık gösterdiği, çoktan seçmeli testin açık uçlu teste göre daha ayırt edici bir test olduğu tespit edilmiştir. Testlerin ayırıcılık düzeyleri arasındaki farkın 0.30 standart sapma kadar olduğu ve test türü değişkeninin test ayırt ediciliği üzerinde küçük düzeyde etkiye sahip olduğu tespit edilmiştir.

Bilişsel alan basamakları açısından incelendiğinde anlama basamağındaki ortalama ayırıcılık düzeyinin test türüne göre anlamlı bir farklılık göstermediği tespit edilmiştir. Bununla birlikte 1, 2, 4, 6 ve 10 no'lu maddelerin anlamlı bir şekilde açık uçlu testte, 2, 8,14 ve 20 no'lu maddelerin ise çoktan seçmeli testte daha ayırıcı olduğu tespit edilmiştir. Uygulama basamağındaki ortalama ayırıcılık düzeyinin test türüne göre anlamlı bir farklılık göstermediği tespit edilmiştir. Bununla birlikte 22 ve 9 no'lu maddelerin açık uçlu testte, 7, 11, 24 ve 26 no'lu maddelerin çoktan seçmeli testte anlamlı bir şekilde daha ayırıcı olduğu tespit edilmiştir. Analiz basamağındaki ortalama ayırıcılık düzeyinin test türüne göre anlamlı farklılık gösterdiği, çoktan seçmeli testin açık uçlu teste göre daha ayırıcı olduğu tespit edilmiştir. Test türünün analiz basamağındaki ortalama ayırıcılık düzeyi üzerinde orta düzeyde bir etkisinin olduğu tespit edilmiştir. 13, 16, 17, 18 ve 27 no'lu maddelerin çoktan seçmeli testte, 12 no'lu maddenin ise açık uçlu testte anlamlı bir şekilde daha ayırıcı olduğu tespit edilmiştir.

Tablo 11: Sosyal Bilgiler Dersi Açık uçlu Test ve Çoktan Seçmeli Test Güvenirlik Katsayılarının Karşılaştırılması

Testler	N	k	Cronbach Alpha /KR 20	W	Cohen d
Açık Uçlu Test	102	27	.747	.940	.136
Çoktan Seçmeli Test	102	27	.731		

p<.05

Tablo 11'e göre açık uçlu test ile çoktan seçmeli testin güvenirlik katsayıları arasında anlamlı bir fark bulunmamıştır ($W < F(101,101) = 1.38$). Buna göre test türü değişkeninin güvenirlik düzeyi üzerinde etkili bir değişken olmadığı, her iki testin de benzer güvenirlikte ölçüm yaptığı söylenebilir.

Tablo 12: Sosyal bilgiler dersi açık uçlu test ve çoktan seçmeli test puanlarının karşılaştırılmasına ilişkin bağımlı gruplar t testi sonuçları

Testler	N	Ortalama	Std. Sapma	sd	t	p
Açık Uçlu Test	102	56.9	97.1			
Çoktan Seçmeli Test	102	69.8	88.1	101	-1.869	.065

p<.05

Tablo 12 incelendiğinde öğrencilerin açık uçlu test puanları ile çoktan seçmeli test puanları arasında istatistiki açıdan anlamlı bir farklılığın olmadığı tespit edilmiştir ($t=-1.869$, $p>.05$). İstatiksel açıdan anlamlı olmamakla birlikte öğrencilerin çoktan seçmeli test puan ortalamasının açık uçlu test puan ortalamasından daha yüksek olduğu görülmektedir.

Tablo 13: Sosyal Bilgiler dersi açık uçlu test ve çoktan seçmeli test ölçümlerinin anlama, uygulama ve analiz düzeyindeki puanlarının karşılaştırılmasına ilişkin bağımlı gruplar t testi sonuçları

	Testler	N	Ortalama	Std. Sapma	sd	t	p
Anlama Düzeyi	Açık Uçlu Test	102	53.1	48.6			
	Çoktan Seçmeli Test	102	50	42.2	101	.741	.460
Uygulama Düzeyi	Açık Uçlu Test	102	51.9	33			
	Çoktan Seçmeli Test	102	50	32.4	101	.489	.626
Analiz Düzeyi	Açık Uçlu Test	102	51.7	41.7			
	Çoktan Seçmeli Test	102	50	36.6	101	.459	.647

p<.05

Tablo 13'e göre öğrencilerin anlama düzeyindeki açık uçlu ve çoktan seçmeli test puanları arasında anlamlı farklılığın olmadığı anlaşılmaktadır ($t=.741$, $p>.05$). Benzer bir bulgu olarak uygulama düzeyindeki açık uçlu ve çoktan seçmeli test puanları arasında anlamlı farklılığın olmadığı tespit edilmiştir ($t=.489$, $p>.05$). Analiz düzeyindeki açık uçlu ve çoktan seçmeli test puanlarının arasında anlamlı farklılığın olmadığı sonucuna ulaşılmıştır ($t=.459$, $p>.05$).

Tablo 14: Sosyal Bilgiler dersi açık uçlu test ve çoktan seçmeli test ölçümlerinin anlama, uygulama ve analiz basamağındaki puanları arasındaki ilişkinin incelenmesine yönelik Pearson korelasyon testi sonuçları

Testler	r	p	r ²
Anlama Basamağı (Açık Uçlu * Çoktan Seçmeli)	.552**	.000	.30
Uygulama Basamağı (Açık Uçlu * Çoktan Seçmeli)	.254**	.010	.06
Analiz Basamağı (Açık Uçlu * Çoktan Seçmeli)	.532**	.000	.28

**p<.01

Tablo 14 açık uçlu ve çoktan seçmeli testlerin anlama basamağındaki puanları arasında pozitif yönde orta düzeyde bir ilişki olduğunu göstermektedir ($r=.552$, $p<.01$, $r^2=.30$). Uygulama basamağındaki puanlar arasında ise pozitif yönde zayıf düzeyde bir ilişki olduğu tespit edilmiştir ($r=.254$, $p<.01$, $r^2=.06$). Analiz basamağındaki puanlar arasında pozitif yönde orta düzeyde bir ilişki olduğu bulgusuna ulaşılmıştır ($r=.532$, $p<.01$, $r^2=.28$).

Tablo 15: Sosyal Bilgiler dersi açık uçlu ve çoktan seçmeli test puanlarının cinsiyete göre karşılaştırılmasına ilişkin iki yönlü ANOVA testi sonuçları

Varyansın Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p
Test Türü (Açık Uçlu-Çoktan Seçmeli)	8534.457	1	8534.457	3.467	.066
Test Türü* Cinsiyet	27.782	1	27.782	.011	.916

p<.05

Tablo 15'e göre öğrencilerin açık uçlu ve çoktan seçmeli test puanlarının cinsiyet değişkeni açısından anlamlı farklılık göstermediği tespit edilmiştir ($F(1, 8534.457)= 3.467$, $p>.05$).

Tablo 16: Sosyal Bilgiler dersi anlama, uygulama ve analiz düzeylerine ilişkin açık uçlu ve çoktan seçmeli test puanlarının cinsiyete göre karşılaştırılmasına ilişkin iki yönlü ANOVA testi sonuçları

	Varyansın Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	F	p
Anlama Düzeyi	Test Türü (Açık Uçlu-Çoktan Seçmeli)	578.950	1	578.950	.617	.434
	Cinsiyet	21172.588	1	21172.588	6.987	.010*
	Test Türü* Cinsiyet	1165.359	1	1165.359	1.242	.268
Uygulama Düzeyi	Test Türü (Açık Uçlu-Çoktan Seçmeli)	204.600	1	204.600	.254	.616
	Cinsiyet	3436.926	1	3436.926	2.592	.111
	Test Türü* Cinsiyet	141.335	1	141.335	.175	.676
Analiz Düzeyi	Test Türü (Açık Uçlu-Çoktan Seçmeli)	132.837	1	132.837	.181	.671
	Cinsiyet	23410.492	1	23410.492	10.883	.001*
	Test Türü* Cinsiyet	499.241	1	499.241	.682	.411

*p<.05

Tablo 16 incelendiğinde anlama düzeyindeki test türü ana etkisinin anlamlı olmadığı ($F(1,578.950)=.617, p>.05$); cinsiyet ana etkisinin anlamlı olduğu ($F(1,21172.588)=6.987, p<.05$) ve test türü *cinsiyet etkileşim etkisinin ise anlamlı olmadığı tespit edilmiştir ($F(1,1165.359)=1.242, p>.05$). Buna göre kız ve erkek öğrencilerin anlama düzeyindeki açık uçlu ve çoktan seçmeli test puanları arasında anlamlı farklılık olmadığı söylenebilir.

Uygulama düzeyinde ise test türü ana etkisinin ($F(1,204.600)=.254, p>.05$), cinsiyet ana etkisinin ($F(1,3436.926)=2.592, p>.05$) ve test türü * cinsiyet etkileşim etkisinin ($F(1,141.335)=.175, p>.05$) anlamlı olmadığı tespit edilmiştir. Buna göre kız ve erkek öğrencilerin uygulama düzeyindeki açık uçlu ve çoktan seçmeli test puanları arasında anlamlı farklılık olmadığı söylenebilir.

Analiz düzeyinde test türü ana etkisinin anlamlı olmadığı ($F(1,132.837)=.181, p>.05$), cinsiyet ana etkisinin anlamlı olduğu ($F(1,23410.492)=10.883, p<.05$) ve test türü * cinsiyet etkileşim etkisinin anlamlı olmadığı ($F(1,499.241)=.682, p>.05$) tespit edilmiştir. Buna göre kız ve erkek öğrencilerin analiz düzeyindeki açık uçlu ve çoktan seçmeli test puanları arasında anlamlı farklılık olmadığı söylenebilir.

4. TARTIŞMA ve SONUÇ

Araştırma sonuçları açık uçlu testler ile çoktan seçmeli testler arasında madde güçlük düzeyleri açısından anlamlı farklılık olduğunu göstermektedir. Buna göre fen bilimleri ve sosyal bilgiler derslerine yönelik çoktan seçmeli testlerin açık uçlu testlerden daha kolay olduğu, test türü değişkeninin bu farklılıkta küçük bir etkiye sahip olduğu tespit edilmiştir. Bilişsel alan basamakları açısından incelendiğinde Fen Bilimleri dersinde uygulama basamağında anlamlı farklılığın olduğu, açık uçlu testin orta güçlükte, çoktan seçmeli testin ise kolay olduğu sonucuna ulaşılmıştır. Uygulama basamağındaki farklılık, bu basamakta yer alan maddelerin disiplinler arası becerileri işe koyuyor olmasından kaynaklanmış olabilir. Nitekim madde bazında incelendiğinde, uygulama basamağında anlamlı farklılık gösteren maddelerin Fen Bilimleri dersine ilişkin bilginin yanı sıra Dünya'nın dönme ve dolanma hareketlerinin sonuçlarını içeren matematiksel işlem gerektirdiği tespit edilmiştir. Literatürde öğrencilerin matematiksel işlemlere yönelik olumsuz tutumları ve hata yapma korkusu sebebiyle matematiksel işlemlerden kaçındıkları ve başarısız oldukları belirtilmektedir (Peker ve Mirasyedioğlu, 2003; Duru, Akgün ve Özdemir, 2005). Öğrencilerin matematiksel işlemlere ilişkin tutum ve becerileri açık uçlu testteki maddeleri yanıtlarken zorlanmalarına sebep olmuş olabilir. Bununla birlikte uygulama basamağındaki maddelerin çoktan seçmeli testte öğrencilere

daha kolay geldiği tespit edilmiştir. Bu durum çoktan seçmeli testlerin öğrencinin doğru cevabı bilmediği halde seçeneklerden yola çıkarak tahmin yoluyla ya da yanlışlıkla doğru yanıt bulmasına imkan veren sınırlılığın kaynaklanmış olabilir.

Sosyal Bilgiler dersine yönelik testler bilişsel alan basamakları açısından incelendiğinde anlama, uygulama ve analiz basamaklarında test türüne göre anlamlı farklılığın olduğu tespit edilmiştir. Üç bilişsel alan basamağında çoktan seçmeli testlerin açık uçlu teste göre daha kolay olduğu sonucuna ulaşılmıştır. Bu durum açık uçlu maddelerin bilgiyi hatırlamayı ve yazılı bir şekilde ifade etmeyi gerektiriyor olmasından kaynaklanabilir. Duran ve Tufan (2017) bilgilerin sınav esnasında hatırlayamama, yanıtları yazılı olarak ifade etmekte zorlanma açık uçlu sorularla yapılan sınavlarda öğrencilerin daha düşük puan almalarına sebep olabileceğini ifade etmektedir. Bunun yanı sıra öğrencilerde açık uçlu soruların zor olduğuna ilişkin ön kabul olabilir ve bu durum öğrencilerin performansını etkilemiş olabilir. Koretz, Lewis, Skewes-Cox ve Burstein (1993) öğrencilerin açık uçlu soruları zor sorular olarak algıladıklarını ve bu soruları yanıtlamadan atlamayı tercih ettiklerini belirtmektedir. Nitekim çoktan seçmeli test kültürünün ağırlıklı olduğu ülkemizde açık uçlu soruların öğrenciler tarafından daha zor algılanması anlaşılabilir. Duran ve Tufan (2017) tarafından yapılan çalışmada öğrencilerin açık uçlu sorulara göre çoktan seçmeli testleri tercih ettikleri bulgusu bu yorumu destekler niteliktedir. Karaca (2015) tarafından çoktan seçmeli test ile kısa cevaplı testin psikometrik özelliklerinin karşılaştırıldığı çalışmada, çoktan seçmeli teste ilişkin madde güçlük değerlerinin kısa cevaplı teste göre daha yüksek olduğu yönündedir. Kan ve Kayapınar (2010) tarafından yapılan çalışmada da benzer bir sonuç elde edilmiştir. Bununla birlikte araştırma bulgusundan farklı olarak Morgil ve Yılmaz (2001) tarafından yapılan çalışmada kısa cevaplı teste ilişkin madde güçlük ortalamasının, çoktan seçmeli testten daha yüksek olduğu ancak; bu bulgunun istatistiksel olarak anlamlı olmadığı belirtilmiştir.

Araştırma sonuçları açık uçlu testler ile çoktan seçmeli testler arasında madde ayırıcılık düzeyleri açısından anlamlı farklılık olduğunu göstermektedir. Buna göre fen bilimleri ve sosyal bilgiler derslerine yönelik çoktan seçmeli testlerin açık uçlu testlerle karşılaştırıldığında daha ayırt edici oldukları, test türü değişkeninin fen bilimleri dersi açısından oluşan farklılıkta büyük etkiye, sosyal bilgiler dersi açısından ise küçük etkiye sahip olduğu tespit edilmiştir. Bilişsel alan basamakları açısından incelendiğinde Fen Bilimleri dersi anlama, uygulama ve analiz basamaklarında açık uçlu ve çoktan seçmeli testlerin madde ayırıcılık düzeyleri arasında anlamlı farklılığın olduğu tespit edilmiştir. Çoktan seçmeli testin üç bilişsel alan basamağında açık uçlu teste göre daha ayırıcı olduğu sonucuna ulaşılmıştır. Sosyal Bilgiler dersinde ise yalnızca analiz basamağında ortalama ayırıcılık düzeyleri açısından çoktan seçmeli test lehine anlamlı bir farklılığa rastlanmıştır. Ortalama ayırıcılık düzeyine ilişkin çoktan seçmeli testler lehine yönelik bulgular, çoktan seçmeli testlerdeki çeldiricilerin iyi işlediğini, eksik ve az bilen ile tam bilen ayırt ettiği şeklinde yorumlanabilir. Kan ve Kayapınar (2010) çeldirici seçenekler iyi oluşturulabildiği takdirde madde ayırıcılık düzeyi açısından çoktan seçmeli testlerin diğer madde tipleriyle oluşturulan testlere göre daha avantajlı olduğunu belirtmektedir. Taib ve Yusoff (2014) da yaptıkları çalışmada çoktan seçmeli maddelerden oluşan testin madde ayırıcılık düzeyi ortalamasının açık uçlu maddelerden oluşan teste göre daha yüksek olduğunu tespit etmiştir. Araştırma bulgusundan farklı olarak Lee, Liu ve Linn (2011) yaptıkları çalışmada açıklama gerektiren maddelerden oluşan testin ayırıcılık gücünün çoktan seçmeli testten daha yüksek olduğunu tespit etmişlerdir.

Araştırma sonuçları fen bilimleri ve sosyal bilgiler derslerine ait açık uçlu ve çoktan seçmeli testlerin güvenilirlik katsayıları arasında istatistiksel açıdan anlamlı bir farklılık olmadığını göstermektedir. Bu bulgu, test türü değişkeninin güvenilir ölçüm yapma düzeyi üzerinde anlamlı farklılık oluşturabilecek bir değişken olmadığı şeklinde yorumlanabilir. Literatürde esnek yanıtlamaya elverişli olmasının, anlatımın duruluğu, yazının güzel, kâğıdın düzenli olması gibi değişkenlerin puanlama yanlılığına sebep olmasının açık uçlu testlerin

güvenirliğini düşürdüğü ifade edilmektedir (Bektaş ve Kudubeş, 2014). Çoktan seçmeli testlerde ise öğrencinin giyinişi, hal ve gidişi, yazı güzelliği, iyi kompozisyon yazma gibi, ölçülen özellik ya da özelliklerin dışındaki etkenlerin puanlamada etkisinin olmadığı belirtilmektedir (Tekin, 2016). Bu noktada testlerin güvenilirlik düzeylerinin denkliğine ilişkin araştırma bulgusu, açık uçlu testlerin puanlanmasında kullanılan dereceli puanlama anahtarlarının testin güvenilirlik düzeyini artırarak, puanlama yanlılığının önüne geçmesinden kaynaklanmış olabilir. Nitekim Ozuru, Briner, Kurby ve McNamara (2013) ve Ko (2010) da yaptıkları çalışmalarda açık uçlu ve çoktan seçmeli testlere ilişkin güvenilirlik kat sayılarının çok yakın olduğunu tespit etmiştir. Literatürde yer alan benzer çalışmaların bulguları da farklı test türlerinin güvenilirlik katsayıları arasında anlamlı farkın olmadığı yönündedir (Özalp, 1992; Karaca, 1996; Uğurlu, 1996; Morgil ve Yılmaz, 2001; Karaca, 2015). Mevcut araştırma bulgusundan farklı olarak çoktan seçmeli testle farklı test türüne ait güvenilirlik değerinin karşılaştırıldığı çalışmalarda çoktan seçmeli test aleyhine anlamlı fark olduğunu gösteren çalışmalar da bulunmaktadır (Kan ve Kayapınar, 2010; Büyükturan ve Demirtaşlı, 2013; Thawabieh, 2016).

Araştırma sonuçları öğrencilerin Fen Bilimleri dersine ait açık uçlu ve çoktan seçmeli test puanları arasında çoktan seçmeli test lehine anlamlı bir farklılık olduğunu göstermektedir. Test türü değişkeninin ise öğrenci puan farkı üzerinde küçük bir etkiye sahip olduğu tespit edilmiştir. Bununla birlikte Sosyal Bilgiler dersine ait açık uçlu ve çoktan seçmeli test puanları arasında anlamlı bir farklılığa rastlanmamıştır. İstatistiksel açıdan anlamlı olmamakla birlikte öğrencilerin çoktan seçmeli test puanlarının açık uçlu test puanlarından daha yüksek olduğu tespit edilmiştir. Çoktan seçmeli testler lehine elde edilen araştırma sonuçları, çoktan seçmeli testlerde seçeneklerin hatırlatıcı olma ihtimali, şans faktörü ve örneklem grubunun çoktan seçmeli testlere ilişkin alışkanlık ve hazır bulunuşluklarından kaynaklanıyor olabilir. Çoktan seçmeli testlerdeki ipucu etkisinin yüksek puana sebep olabileceği ifade edilmektedir (Schuwirth, Vleuten ve Donkers; 1996). Braun, Bennett, Frye ve Soloway (1990) çoktan seçmeli maddelerde öğrencilerin eleme yöntemiyle doğru cevaba ulaşabileceğini ancak açık uçlu maddelerde böyle bir durum söz konusu olmadığını belirtmektedir.

Literatürde araştırma bulgusuyla örtüşen çalışmalara rastlanmaktadır. Vasan ve arkadaşları (2017) tarafından yapılan çalışmada öğrencilerin çoktan seçmeli test puanlarının açık uçlu test puanlarına göre daha yüksek olduğu tespit edilmiştir. Temizkan ve Sallabaş (2011) tarafından yapılan çalışmada da öğrencilerin okuduğunu anlama düzeyi açısından çoktan seçmeli maddelerde açık uçlu maddelere göre daha başarılı oldukları tespit edilmiştir. Ko (2010) tarafından yapılan çalışmada öğrencilerin çoktan seçmeli test ve açık uçlu test puanları arasında çoktan seçmeli test lehine anlamlı farklılık olduğu tespit edilmiştir. Farklı formatlardaki test türlerinin karşılaştırıldığı ve çoktan seçmeli testler lehine bulguları olan başka çalışmalara da rastlamak mümkündür (Thawabieh, 2016; Danili ve Reid, 2005). Buna göre araştırma bulgusunun literatürdeki diğer çalışmalarla paralellik gösterdiği söylenebilir. Araştırma bulgusundan farklı olarak Kan ve Kayapınar (2010) tarafından yapılan çalışmada çoktan seçmeli test ile kısa cevaplı testten elde edilen puanlar arasında anlamlı bir farklılığa rastlanmamıştır. Benzer bir bulgu da Morgil ve Yılmaz (2001) tarafından yapılan çalışmada elde edilmiştir.

Araştırma sonuçları Fen Bilimleri ve Sosyal Bilgiler derslerinde anlama, uygulama ve analiz basamaklarında açık uçlu ve çoktan seçmeli test puanları arasında anlamlı farklılık olmadığını göstermektedir. Bu bulgu öğrencilerin bilişsel alan basamaklarındaki performanslarının madde türüne göre değişiklik göstermediği şeklinde yorumlanabilir. Bunun yanı sıra araştırma sonuçları Fen Bilimleri ve Sosyal Bilgiler derslerine yönelik açık uçlu ve çoktan seçmeli testlerin anlama, uygulama ve analiz basamaklarında pozitif yönde anlamlı ilişki olduğunu göstermektedir. Buna göre öğrencilerin açık uçlu teste ait bilişsel alan basamaklarındaki puanları arttıkça, çoktan seçmeli teste ait bilişsel alan basamaklarındaki puanlarının da arttığı söylenebilir. Başka bir ifade ile öğrencilerin açık uçlu teste ait bilişsel alan

basamaklarındaki puanları azaldıkça, çoktan seçmeli teste ait bilişsel alan basamaklarındaki puanlarının da azaldığı söylenebilir. Bu bulgu bilişsel alan basamaklarındaki ölçümlerin test türü açısından farklılaşmadığı, benzer düzeyde ölçüm yapıldığı şeklinde yorumlanabilir. Araştırma bulgusuyla benzer olarak Hancock (1994) tarafından yapılan çalışmada çoktan seçmeli ve açık uçlu testin bilgi, anlama, uygulama ve analiz basamaklarındaki ölçümleri arasında yüksek düzeyde ilişki olduğu, farklı formatlardaki testlerin aynı taksonomik düzeyde ölçüm yaptığı tespit edilmiştir. Araştırma bulgusunun literatürdeki diğer araştırmaların bulgusuyla örtüştüğü söylenebilir (Traub, 1993; Simkin ve Kuechler, 2005).

Araştırma sonuçları Fen Bilimleri ve Sosyal Bilgiler derslerine yönelik açık uçlu ve çoktan seçmeli testlere ilişkin puanlar arasında cinsiyet değişkeni açısından anlamlı bir farklılığın olmadığını göstermektedir. Bilişsel alan basamakları açısından değerlendirildiğinde ise Sosyal Bilgiler dersine ait açık uçlu ve çoktan seçmeli testlerin anlama, uygulama ve analiz basamaklarına ilişkin puanları arasında cinsiyet değişkeni açısından anlamlı bir farklılık olmadığı tespit edilmiştir. Morgil ve Yılmaz (2001) tarafından yapılan çalışmada farklı madde türündeki öğrenci başarısının cinsiyete göre değişmediği bulgusu elde edilmiştir. Bununla birlikte Fen Bilimleri dersi açık uçlu ve çoktan seçmeli test puanlarının analiz basamağında cinsiyet değişkeni açısından farklılık gösterdiği tespit edilmiştir. Buna göre kız öğrencilerin analiz basamağında açık uçlu testte, erkek öğrencilerin ise çoktan seçmeli testte daha başarılı oldukları tespit edilmiştir. Benzer olarak Wright vd., (2016) tarafından yapılan çalışmada farklı test türlerine ait düşük düzeydeki bilişsel basamak puanları arasında cinsiyet açısından anlamlı bir farklılığa rastlanmazken, yüksek düzeydeki bilişsel alan basamak puanları arasında cinsiyet açısından anlamlı farklılık olduğu tespit edilmiştir. Leaver ve van Walbeek (2006) yaptıkları çalışmada kız öğrencilerin analiz basamağında çoktan seçmeli testlerde erkeklerden daha düşük puan aldığı bulgusuna ulaşmışlardır. Du Plessis ve Du Plessis (2009) yaptığı çalışmada çoktan seçmeli testte başarılı olan erkek öğrencilerin, aynı soruların açık uçlu versiyonlarında başarısız olduklarını tespit etmiştir. Nitekim literatürde kız öğrencilerin açık uçlu testlerde, erkek öğrencilerin ise çoktan seçmeli testlerde daha başarılı olduğunu gösteren çalışmalar yer almaktadır (Bolger ve Kellaghan, 1990).

Araştırma sonuçları aynı üniteye yönelik olarak geliştirilen farklı formatlardaki testler arasında madde güçlük, madde ayırıcılık açısından fark olduğunu göstermektedir. Fen Bilimleri ve Sosyal Bilgiler derslerine ait açık uçlu ve çoktan seçmeli testlerin güvenilirlik katsayıları arasında istatistiksel açıdan anlamlı bir farklılık olmadığı tespit edilmiştir. Öğrencilerin Fen Bilimleri dersine yönelik çoktan seçmeli test puanlarının açık uçlu test puanlarından daha yüksek olduğu tespit edilmiştir. Sosyal Bilgiler dersinde öğrenci puanları arasında ise test türü açısından anlamlı bir farklılığın olmadığı tespit edilmiştir. Fen Bilimleri ve Sosyal Bilgiler derslerinde öğrencilerin açık uçlu ve çoktan seçmeli testlere ilişkin bilişsel alan basamaklarındaki puanları arasında anlamlı bir farklılığın olmadığı, farklı test türündeki bilişsel alan basamakları arasında anlamlı ilişki olduğu tespit edilmiştir. Öğrencilerin her iki derse yönelik açık uçlu ve çoktan seçmeli puanlarının cinsiyet değişkeni açısından anlamlı farklılık göstermediği sonucuna ulaşılmıştır. Sosyal Bilgiler dersinde bilişsel alan basamaklarındaki puanların cinsiyet açısından farklılaşmadığı, Fen Bilimleri dersinde ise analiz basamağında cinsiyet açısından anlamlı farklılık olduğu sonucuna ulaşılmıştır.

Araştırma sonuçları açık uçlu test ile çoktan seçmeli testin psikometrik özelliklerinin ve öğrenci performansı açısından karşılaştırılmasına dayanmaktadır. Eşleştirme, doğru-yanlış, boşluk doldurma gibi farklı test formatlarının karşılaştırılmasına ilişkin çalışmalar yapılabilir. Grafik, resim vb. eklendiği farklı test formatları karşılaştırılabilir. Testlerin madde güçlük ve ayırıcılık düzeylerine ilişkin bulgular, öğrencilerle yapılan görüşmelerle derinleştirilebilir. Bilişsel alan basamaklarının karşılaştırıldığı farklı madde türlerini içeren çalışmalar yapılabilir.

5. KAYNAKLAR

- Akay, H., Soybaş, D. ve Argün, Z. (2006). Problem kurma deneyimleri ve matematik öğretiminde açık-uçlu soruların kullanımı. *Kastamonu Eğitim Dergisi*, 14(1), 129–146.
- Badger, E. ve Thomas, B. (1992). Open-ended questions in reading. *Practical Assessment, Research & Evaluation*, 3(4), 03.
- Bahar, M., Nartgün, Z., Durmuş, S. ve Bıçak, B. (2012). *Geleneksel-Tamamlayıcı Ölçme ve Değerlendirme Teknikleri: Öğretmen El Kitabı*. Pegem Akademi.
- Baştürk, S. (2014). Çoktan seçmeli testler. İçinde *Eğitimde Ölçme ve Değerlendirme*. Nobel Akademik Yayıncılık.
- Bektaş, M. ve Kudubeş, A. A. (2014). Bir Ölçme ve Değerlendirme Aracı Olarak: Yazılı Sınavlar. *Dokuz Eylül Üniversitesi Hemşirelik Fakültesi Elektronik Dergisi*, 7(4).
- Beller, M. ve Gafni, N. (2000). Can item format (multiple choice vs. open-ended) account for gender differences in mathematics achievement? *Sex Roles*, 42(1–2), 1–21.
- Birgili, B. (2014). *Open Ended Questions As An Alternative To Multiple Choice: Dilemma In Turkish Examination System*. Middle East Technical University.
- Bolger, N. ve Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. *Journal of Educational Measurement*, 27(2), 165–174.
- Braun, H. I., Bennett, R. E., Frye, D. ve Soloway, E. (1990). Scoring constructed responses using expert systems. *Journal of Educational Measurement*, 27(2), 93–108.
- Bush, M. (2001). A multiple choice test that rewards partial knowledge. *Journal of Further and Higher Education*, 25(2), 157–163.
- Büyüköztürk, Ş., Çokluk, Ö. ve Köklü, N. (2010). *Sosyal Bilimler İçin İstatistik* (5. Baskı). Ankara: Pegem Akademi.
- Büyükturan, E. B. ve Demirtaşlı, N. Ç. (2013). Comparing the psychometric characteristics of multiple choice tests and structural communication grids. *Eğitim Bilimleri Fakültesi Dergisi*, 46(1), 395.
- Çelik, E. (2015). Psychometric Properties of Brief Sensation Seeking Scale (BSSS-4) in a Sample of Turkish Adolescents. *Sakarya University Journal of Education*, 5(3), 42–51.
- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (Second Edition). Lawrence Erlbaum Associates.
- Cooney, T. J., Sanchez, W. B., Leatham, K. ve Mewborn, D. S. (2004). *Open-ended assessment in math: A searchable collection of 450+ questions*. Heinemann. Quotations taken from online version, books. heinemann.com/math.
- Danili, E. ve Reid, N. (2005). Assessment formats: do they make a difference? *Chemistry Education Research and Practice*, 6(4), 204–212.
- Demirbilek, S. (2015). *Öğretmen Adaylarının Eğitimde Ölçme Ve Değerlendirme Dersindeki Kavram Yanılgılarının İncelenmesi* (Yüksek Lisans). Hacettepe Üniversitesi.
- Dochy, F., Janssens, S. ve Struyven, K. (2005). Students' perceptions about evaluation and assessment in higher education: a review 1. *Assessment & Evaluation in Higher Education*, 30(4), 325–341.
- Du Plessis, S. ve Du Plessis, S. (2009). *A new and direct test of the 'gender bias' in multiple-choice questions*.
- Duran, E. ve Tufan, B. S. (2017). The Effect Of Open-Ended Questions And Multiple Choice Questions On Comprehension, *International Journal of Languages' Education and Teaching*, 5(1), 242–254.
- Duru, A., Akgün, L. ve Özdemir, M. E. (2005). İlköğretim Öğretmen Adaylarının Matematiğe Yönelik Tutumlarının İncelenmesi. *Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi*, (11).
- Geer, J. G. (1988). What do open-ended questions measure? *Public Opinion Quarterly*, 52(3), 365–367.
- Güler, N. (2017). *Eğitimde Ölçme ve Değerlendirme* (10.Baskı). Ankara: Pegem Akademi.
- Gültekin, S. (2014). Performans dayanaklı değerlendirme. İçinde R. N. Demirtaşlı (Ed.), *Eğitimde ölçme ve değerlendirme*. Ankara: Edge Akademi.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E. ve Tatham, R. L. (2006). *Multivariate data analysis 6th ed. Uppersaddle River: Pearson Prentice Hall*.
- Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *The Journal of Experimental Education*, 62(2), 143–157.
- Kan, A. ve Kayapınar, U. (2010). Yabancı Dil Eğitiminde Aynı Davranışları Yoklayan Çoktan Seçmeli ve Kısa Cevaplı İki Testin Madde ve Test Özelliklerinin Karşılaştırılması. *Eğitim Ve Bilim*, 32(142).

Karaca, E. (1996). *Seçme Gerektiren, Kısa Cevaplı ve Doğru-Yanlış Testlerinin Madde ve Test Özelliklerinin Karşılaştırılması*. Yayınlanmamış yüksek lisans tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.

Karamustafaoğlu, S. ve Tutar, M. (2016). 6. Sınıf Dünya'mız, Ay ve Yaşam Kaynağımız Güneş Ünitesi'ne yönelik bir başarı testi geliştirme. *Pegem Atf İndeksi*, 303–320.

Karataş, F. Ö., Köse, S., ve Coştu, B. (2003). Öğrenci Yanılgılarını Ve Anlama Düzeylerini Belirlemede Kullanılan İki Aşamalı Testler. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 13(13), 54–69.

Kılıç, A. F. ve Kelecioğlu, H. (2016). TEOG Ortak ve Mazeret Sınavındaki Türkçe ve Matematik Alt Testlerinin Psikometrik Özelliklerinin Karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 7(1).

Klufa, J. (2015). Multiple Choice Question Tests—Advantages and Disadvantages.

Ko, M. H. (2010). A Comparison of Reading Comprehension Tests: Multiple-Choice vs. Open-Ended. *English Teaching*, 65(1).

Koretz, D., Lewis, E., Skewes-Cox, T. ve Burstein, L. (1993). Omitted and Not-Reached Items in Mathematics in the 1990. *National Assessment of Educational Progress*.

Krippendorff, K. (2007). Computing Krippendorff's alpha reliability. *Departmental Papers (ASC)*, 43.

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

Leaver, R. ve van Walbeek, C. (2006). “Gender bias” in Multiple-Choice Questions: DoES the type of question make a difference?

Lee, H.-S., Liu, O. L. ve Linn, M. C. (2011). Validating Measurement of Knowledge Integration in Science Using Multiple-Choice and Explanation Items. *Applied Measurement in Education*, 24(2), 115–136. <https://doi.org/10.1080/08957347.2011.554604>

Lenhard, W. ve Lenhard, A. (2016). Calculation of Effect Sizes. *Psychometrica*. <https://doi.org/10.13140/RG.2.1.3478.4245>

Morgil, F. İ., ve Yılmaz, A. (2001). Kimya eğitiminde farklı madde türlerinin psikometrik özellikleri ve öğrenci başarıları bakımından karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20(20).

Organisation for Economic Cooperation and Development (OECD) (2007). *PISA 2006 Science Competencies for Tomorrow's World*. Retrieved February 17, 2018, from <http://www.pisa.oecd.org>.

Özalp, D. (1992). Kısa Cevaplı ve Çoktan Seçmeli Maddelerden Oluşan Testlerin Psikometrik Özelliklerinin Karşılıklı Olarak İncelenmesi. *HÜ Sosyal Bilimler Enstitüsü, Yüksek Lisans Tezi, Ankara*.

Ozuru, Y., Briner, S., Kurby, C. A. ve McNamara, D. S. (2013). Comparing comprehension measured by multiple-choice and open-ended questions. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, 67(3), 215.

Peker, M. ve Mirasyedioğlu, Ş. (2003). Lise 2. Sınıf Öğrencilerinin Matematik Dersine. Yönelik Tutumları ve Başarıları Arasındaki İlişki *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 14(14), 157–166.

PISA. (2015). *Uluslararası Öğrenci Değerlendirme Programı PISA 2015 Ulusal Raporu*. M.E.B. Ölçme Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü.

Quaigrain, K. ve Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, 4(1), 1301013.

Rauch, D. P. ve Hartig, J. (2010). Multiple-choice versus open-ended response formats of reading test items: A two-dimensional IRT analysis. *Psychological Test and Assessment Modeling*, 52(4), 354–379.

Ringim, K. J., Razalli, M. R. ve Hasnan, N. (2012). A framework of business process re-engineering factors and organizational performance of Nigerian banks. *Asian Social Science*, 8(4), 203.

Simkin, M. G. ve Kuechler, W. L. (2005). Multiple-choice tests and student understanding: what is the connection? *Decision Sciences Journal of Innovative Education*, 3(1), 73–98.

Taib, F. ve Yusoff, M. S. B. (2014). Difficulty index, discrimination index, sensitivity and specificity of long case and multiple choice questions to predict medical students' examination performance. *Journal of Taibah University Medical Sciences*, 9(2), 110–114. <https://doi.org/10.1016/j.jtumed.2013.12.002>

Tekin, H. (2016). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Yargı Yayın Evi.

Tekindal, S. (2000). Klasik Yazılı Sınavla Ve Çok Sorulu Testle Elde Edilen Ölçümlerin Güvenirlik Ve Geçerliği. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 8(8), 38–46.

Temizkan, M., ve Sallabaş, M. E. (2011). Okuduğunu anlama becerisinin değerlendirilmesinde çoktan seçmeli testlerle açık uçlu yazılı yoklamaların karşılaştırılması. *Dumlupınar Üniversitesi Sosyal Bilimler Dergisi*, 30, 207–220.

Thawabieh, A. M. (2016). A Comparison Between Two Test Item Formats: Multiple-Choice Items And Completion Items. *British Journal of Education*, 4(8), 32–43.

Traub, R. E. (1993). On the equivalence of the traits assessed by multiple-choice and constructed-response tests. *Construction versus Choice in Cognitive Measurement: Issues in Constructed Response, Performance Testing, and Portfolio Assessment*, 29–44.

Turgut, M. F. (1983). *Eğitimde Ölçme ve Değerlendirme Metotları* (İkinci Baskı). Ankara: Saydam Matbaacılık.

Turgut, M. F. ve Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme*. Pegem Akademi.

Uğurlu, N. (1996). Aynı Davranışı ölçmeye Yönelik Kısa Cevaplı, Çoktan Seçmeli ve Kompozisyon Tipi Maddelerden Oluşa Testlerin Özelliklerinin İncelenmesi. *Yayımlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.*

Üstüner, A. ve Şengül, M. (2004). Çoktan seçmeli test tekniğinin Türkçe öğretimine olumsuz etkileri. *Fırat Üniversitesi Sosyal Bilimler Dergisi*, 14(2), 197–208.

Vasan, M., Cheryl, A., DeFouw, D. O., Holland, B. K. ve Vasan, N. S. (2017). Analysis of testing with multiple choice versus open-ended questions: Outcome-based observations in an anatomy course. *Anatomical Sciences Education*.

Wilson, F. R., Pan, W. ve Schumsky, D. A. (2012). Recalculation of the Critical Values for Lawshe's Content Validity Ratio. *Measurement and Evaluation in Counseling and Development*, 45(3), 197–210.

Wooten, M. M., Cool, A. M., Prather, E. E. ve Tanner, K. D. (2014). Comparison of performance on multiple-choice questions and open-ended questions in an introductory astronomy laboratory. *Physical Review Special Topics-Physics Education Research*, 10(2), 020103.

Wright, C. D., Eddy, S. L., Wenderoth, M. P., Abshire, E., Blankenbiller, M. ve Brownell, S. E. (2016). Cognitive difficulty and format of exams predicts gender and socioeconomic gaps in exam performance of students in introductory biology courses. *CBE-Life Sciences Education*, 15(2), ar23.

Yaman, S. (2016). Çoktan Seçmeli Madde Tipleri Ve Fen Eğitiminde Kullanılan Örnekleri. *Gazi Eğitim Bilimleri Dergisi*, 2(2).

Yazıcıoğlu, E. B. (2007). *Çoktan Seçmeli Testler ile Yapılandırılmış Gridlerin Psikomotor Özellikler Açısından Karşılaştırılması*. Yayımlanmamış Yüksek Lisans Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.

Extended Abstract

The assessment and evaluation is fundamental components of education process. It is considered that measuring format is important measuring as well. The item formats have different advantages and limitations with regard to base on construct which is intended to measure. Multiple choice questions are widely used in testing due to advantage quick and easy scoring, especially in crowd of students. On the other hand, students can obtain points by guessing the right answers and this fact affects reliability of test scores. Besides, multiple choice tests may remain insufficient to measure higher levels of cognitive domains. Open ended questions require students to create own responses, produce a variety of solutions and process complex thinking skills. It is considered that open ended question tests are useful to measure depth learning. Limitations of open ended questions test are correction of that kind of questions is mostly time consuming and it can be also subjective.

This study aimed to compare open ended questions test and modified test based on multiple choice questions in terms of psychometric features and student performance. Open ended questions test and modified multiple choice test with regard to item difficulty, item discrimination, reliability, gender, student performance scores and cognitive domain levels are understanding, applying and analyzing.

Research data was obtained with open ended question tests were developed towards Science and Social Sciences lessons and modified multiple choice tests based on these open ended questions tests. Research sample consisted of of 102 participants who attend 5th grade secondary school in Samsun, Turkey. The normality of research data was checked with skewness and kurtosis values and z value calculated. Data was in normally distribution pattern regarding with z values were in ± 2.58 range. The differences between tests with regard to item difficulty index and item discrimination index were analyzed with z test for proportions. Feldt test was used to compare internal consistency index of tests. Comparing test formats according to mean were tested with paired sample t test. Correlation between cognitive domain levels was tested with Pearson Correlation test. Comparing difference between test formats according to gender was analyzed with two way ANOVA.

Research results show that average item difficulty indices of multiple choice tests were higher than open ended questions tests for each lesson. In other words multiple choice tests were easier than open

ended questions. It was found that test item format has small effect size on item difficulty. There was significant difference between application cognitive domain levels of each open ended and multiple choice tests regarding item difficulty mean at Science lesson. There were significant difference between item difficulty levels of comprehension, application and analysis in each test format at Social Science lesson. It was found that average item discrimination indices were higher than open ended questions tests for each lesson. It shows that multiple choice tests are more discriminative than open ended questions tests. It could be said that distracters of multiple choice tests has processed well to discriminate miscognizant and smatters. Beside there were significant differences between all cognitive domain levels of tests regarding item discrimination indexes at Science lesson. There was significant difference only between analysis cognitive domain levels of test at Social Sciences lesson. Multiple choice test cognitive domain levels were more discriminative than open ended tests. Research results show that there were no significant differences between multiple choice test and open ended questions test according to internal consistency indices for each lesson.

Research results show that there was significant difference between multiple choice test and open ended questions test score in Science lesson. However there was no significant difference between multiple choice test and open ended questions test score in Social Science lesson. Multiple choice test score was higher than open ended questions test score at Science lessons. Not statistically, multiple choice test score was higher than open ended questions test score at Social Science lesson, too.

There was no significant difference between cognitive domain levels of each tests in both lessons. Beside there were significant positive relationships between cognitive domain levels of each test in both lessons. It means that tests with different format measured same constructs. There was no significant difference between open ended and multiple choice test with regarding to gender variable for both lessons. However there was significant difference in analysis level between boys' and girls' open ended and multiple choice test scores in Science lesson. Girls were more successful at open ended questions than multiple choice test. Boys were more successful at multiple choice test that open ended tests. There was no significant difference between cognitive domain levels of tests regarding with gender in Social Science lesson.

This research shows that test item formats cause differences with regards of item difficulty, item discrimination, student performance and cognitive domain level. This research was conducted base multiple choice test and open ended questions test formats. Further researches may compare true/false, completion test formats.