



**Mustafa Kemal Üniversitesi Eğitim Fakültesi Dergisi**  
**Mustafa Kemal University Journal of the Faculty of Education**  
**Yıl/Year: 2023 ◆ Cilt/Volume: 7 ◆ Sayı/Issue: 12, s. 161-183**

**YENİLENMİŞ BLOOM TAKSONOMİSİNÉ GÖRE “VÜCUDUMUZDAKİ SİSTEMLER” BAŞARI TESTİ GELİŞTİRME ÇALIŞMASI**

**Başak AKAR**

Gaziantep Üniversitesi, Gaziantep Eğitim Fakültesi, [basakakar8989@gmail.com](mailto:basakakar8989@gmail.com)

Orcid: 0009-0002-5798-9705

**İsmet GÜNEŞ**

MEB, Öğretmen, [ismetgunes001@gmail.com](mailto:ismetgunes001@gmail.com)

Orcid: 0009-0008-4310-2062

**Doç. Dr. İbrahim YILDIRIM**

Gaziantep Üniversitesi, Gaziantep Eğitim Fakültesi [iyildirim84@gmail.com](mailto:iyildirim84@gmail.com)

Orcid: 0000-0002-4137-2025

**Özet**

Araştırmanın amacı, Fen Bilimleri dersinin 6.sinif öğretim programına ilişkin “Vücutumuzdaki Sistemler” ünitesiyle ilgili başarı testi geliştirmektir. Bu amaçla belirlenen kazanımlara uygun 50 soruluk madde havuzu oluşturulmuştur. Ön deneme sonrası Fen Bilimleri, Biyoloji, Ölçme-Değerlendirme ve Türkçe eğitiminde uzman görüşlerine başvurularak soru sayısı 26'ya indirilmiştir. Bu test, 2022-2023 eğitim-öğretim yılında Güneydoğu'nun bir ilinin merkez iki ilçesinde eğitime devam eden 157; 6.sinif öğrencisine uygulanmıştır. Geçerliği-güvenirliği incelenen başarı testinin kapsam geçerliği için belirtke tablosu hazırlanmıştır. Öğrencilerin test sorularına verdikleri cevaplar doğrultusunda madde analizi yapılmıştır. Her madde için ayrıt edicilik-madge güclük indeksleri hesaplanmıştır. Madde analizi sonucunda 26 sorudan oluşan “Vücutumuzdaki Sistemler” ünitesi başarı testi oluşturulmuştur. TAP programı ile testin ortalama güçlüğüünün 0,493/orta güçlükte olduğu ayrıt ediciliklerinin ortalamasının 0,527; KR20 güvenirlik değerinin 0,85 olduğu tespit edilmiştir. Eşdeğer yarılar toplam güvenirlik değeri ise 0,83 hesaplanmıştır. Test yapısının geçerliğinin kanıtı; açımlayıcı faktör analiziyle (AFA) sağlanarak FACTOR programının tek faktörlü yapı önerdiğini, toplam varyansın da %5'ini açıkladığı anlaşılmıştır. Geçerlik çalışmalarına ek maddelerin cinsiyet yanılılığı analizleri Mantel-Haenszel (MH) teknikleri ile belirlenmiştir. Yanlık analizi sonucunda tüm maddelerin düşük DMF değerinde olduğu görülmüştür. Sonuç olarak fen bilimleri derslerinde kullanılabilen geçerli-güvenir bir başarı testi geliştirildiği söylenebilir.

**Anahtar Kelimeler:** Başarı testi, Fen Bilimleri, Vücutumuzdaki Sistemler, Geçerlik, Güvenirlik

**DEVELOPING AN ACHIEVEMENT TEST ON “SYSTEMS IN OUR BODY” BASED ON THE REVISED BLOOM’S TAXONOMY**

**Abstract**

The aim of the research was to develop an achievement test for the unit titled "Systems in Our Body" in the sixth grade science curriculum. For this purpose, an item pool with 50 questions was created in compliance with the related learning outcomes. After the pilot test, experts in science education, measurement and assessment, and Turkish language education were consulted, and the number of questions was reduced to 26. This test was applied to 157 sixth-grade students in two central districts of the South-Eastern region in 2022/2023 academic year. The validity and reliability of the achievement test were checked. A table of specifications was created for the content validity. Based on the students' answers to the test items, an item analysis was performed and the indices of distinction and difficulty were calculated for each item separately. As a result of the item analysis, an achievement test consisting of 26 questions for the unit "Systems in our body" was created. The TAP program revealed that the mean difficulty score of the test was 0.493/ distinctiveness score was 0.527; the KR20 reliability score was 0.85. The overall reliability of the equivalent halves was 0.83. Evidence of the validity of the test structure was provided by exploratory factor analysis (EFA). As a result of the EFA, it was found that the FACTOR program suggested a one-component structure and the one-component structure explained 5% of the total variance. In addition to the validity studies, gender bias analyses of the test items were performed, and the DIF values of the items were determined using the Mantel-Haenszel (MH) techniques. After the analysis load, all items were found to have low DIF values. Based on the results obtained, it can be concluded that a valid and reliable performance test has been developed that can be used in science lessons.

**Key Words:** Achievement test, Science, Systems in our Body, Validity, Reliability

## Giriş

Fen Bilimleri, evrendeki olguları birbirine bağlayan unsurları bilimsel yasalar ile açıklayan doğayı ve doğadaki ilişkileri anlamaya yarayan bir bilim dalıdır. Bilim ve teknolojinin hızla gelişmesi, bireylerin ve toplumun değişen gereksinimleri fen bilimleri öğretim programını da doğrudan etkilemiştir. Öğretim programı, eğitim programında yer alan bilgi ve becerileri, o programa uygun olarak planlı bir şekilde kazandırmayı amaçlayan programdır (Yıldırım, 2016). Öğretim programı, hedef, içerik, öğrenme-öğretim süreci ve ölçme değerlendirme olmak üzere dört ögeden oluşmaktadır. Ölçme kavramı; herhangi bir nesnenin belli bir özelliğinin sayı ve semboller ile ifade edilmesi olarak tanımlanır (Yıldız vd., 2019). Ölçme sonuçlarının belirli bir ölçütle kıyaslanarak bir yargıya varma sürecine ise değerlendirme denir (Bahar vd., 2012). Eğitimde ölçme ve değerlendirme, öğrencilerin gelişimlerini izlemek, öğrenme eksikliklerini belirlemek, hazırlıbulunuşluklarını, eğitim-öğretim sürecinde yapılan çalışmaların, kullanılan yöntem, teknik araç ve gereçlerin etkisini belirlemek, programların istenilen hedefe uygunluğunu test etmek ve öğrencilere dönüt vermek için önemlidir (Algın, 2008). Türkiye'de yer alan öğretim programları da bu bağlamda hazırlanmaktadır. Evrendeki olguları birbirine bağlayan, bilimsel yasalar ile açıklayan doğayı ve doğadaki ilişkileri anlamaya yarayan fen bilimleri dersinin öğretim programları da bu kapsamda hazırlanmaktadır. Bilim ve teknolojinin hızla gelişmesi, bireylerin ve toplumun değişen gereksinimleri fen bilimleri öğretim programını da etkilemektedir.

Ülkemizde Fen Bilimleri öğretim programları davranışçı yaklaşımından araştırma ve sorgulamaya dayalı yapılandırmacılığa geçerek 2014 itibarı ile geliştirilmiş en son ise 2018 yılında revize edilmiştir (Gömlekşiz, 2015). Yenilenen öğretim programına göre yapılandırmacı öğrenme kuramı ön plana çıkmıştır (Gömlekşiz, 2005). Öğrencinin fen derslerinde belirlediği hedefleri, görevleri ve sonuçları içselleştirme derecesinin yanı sıra elde ettiği başarıları, geçerli ve güvenilirliğe sahip ölçme araçlarıyla belirlenmelidir (Gönen, Kocakaya ve Kocakaya, 2011). Öğrencilerin başarı seviyesini ölçmek için birçok ölçme aracı geliştirilmiştir. Bunlar çoğunlukla çoktan seçmeli maddeler, doğru yanılış maddeleri, eşleşen maddeler ve kısa cevaplı maddelerdir. Bunlardan en sık kullanılan çoktan seçmeli test maddeleridir (Gönen vd., 2011). Ölçme değerlendirme araçlarından sıkılıkla kullanılan çoktan seçmeli iki bölümden oluşur: ögenin kökü ve seçeneklerdir. Bunların ilki soru cümlesi ve sorunun bir kısmı köküdür; seçenekler ise maddenin doğru cevabını ve celdiricileri, yani yanlış cevapları içeren cümlelerdir (Ogan, 2004). Çalışmada Bloom taksonomisi esas alınarak çoktan seçmeli sorular oluşturulmuştur.

1956 yılında Bilgi, Kavrama, Uygulama, Analiz, Sentez ve Değerlendirme olarak altı basamaktan oluşan ilk taksonomik süreçler Bloom'un çalışmasıyla başlamıştır. Ölçme aracı geliştirilmesinde bu taksonomik sınıflandırmadan uzun yıllar eğitim hedeflerine uygun şekilde yararlanılmıştır (Dinçer ve Yıldırım, 2022).

Bloom Taksonomisinin tek boyutta olduğu, asıl olan sentez düzeyinin değerlendirmeden daha karmaşık olduğu hatta sentezin değerlendirmeyi kapsadığı yönünde eleştiriler olmuştur. Dünyadaki gelişmelere ayak uydurmak, öğretim yöntem ve tekniklerinin farklılaşması, gelişim – öğrenme psikolojisinin dikkate alınması, ölçme – değerlendirme tekniklerinin geliştirilmesi ile yeni bir taksonomiye ihtiyaç duyulmuştur (Bumen, 2010). Anderson vd. (2006), Hatırlama, Anlama, Uygulama, Analiz, Değerlendirme ve Yaratma olarak altı basamaklı Yenilenmiş Bloom

Taksonomisi’ni oluşturmuşlardır. Çalışmada bu taksonomiye göre başarı testi geliştirilmesi amaçlanmıştır.

Yenilenmiş Bloom Taksonomisi ile sınıflandırma işlemi tek boyutlu olmaktan çıkarak iki boyutlu hale gelmiş olması nedeniyle maddelerin bilgi ve eylem halleri birbirinden ayrılarak daha anlaşılır bir hale gelmesi açısından çalışmayı önemli kılmaktadır. Bu sayede iki boyutlu bu taksonomide özellikle bilgi boyutunun dört kategoriye ayrılması öncekine göre daha geniş alanlarda kullanılmasını sağlamıştır (Bekdemir ve Selim, 2008).

Başarı testi geliştirme çalışması hem Bloom hem de Yenilenmiş Bloom taksonomisine göre hazırlanılsa da günümüz eğitim sistemi ve öğrenme süreçlerinin gerektirdiği becerileri tam olarak değerlendirmek için Yenilenmiş Bloom Taksonomisi çalışmada tercih edilmiştir. Ayrıca Yenilenmiş Bloom taksonomisinde öğrencilerin daha üst düzey düşünme becerilerini geliştirmeye yardımcı olması ve öğrencilerin bilgiyi sadece hatırlamayı değil, etkili şekilde kullanabilme becerisini de ölçmesi açısından çalışmada bu taksonomi tercih edilmiştir.

Ölçme aracı geliştirme sürecinde en önemli basamaklardan bir diğeri ise yanlışlık analizleridir. Yanlılık, aynı yetenek düzeyindeki iki ya da daha fazla grubun maddeleri doğru yanıtlama olasılıklarının farklılaşması olarak tanımlanır (Aybek vd., 2021). Ölçme aracının bir grup lehine (cinsiyet, etnik köken, kültür vb.) yanlı davranışları, ölçme aracının geçerliğini doğrudan etkiler. Bu nedenle geliştirilmesi planlanan bir ölçme aracının yanlışlık analizleri yapılmalı ve yanlışlık gösteren maddeler ölçme aracından çıkartılmalıdır. Bu araştırmada Klasik Tepki Kuramı'na dayalı yanlışlık analizi tekniklerinden Mantel-Haenszel Tekniği (MH) kullanılarak yanlışlık analizleri yapılacak ve sonuçları yorumlanacaktır.

Mantel-Haenszel Tekniği (MH) ile kategorik değişkenler arasındaki ilişkiler kullanırken bir üçüncü kontrol değişkeninin yürütülmesine yardımcı olur. Böylece, değişkenler arasındaki ilişkiyi daha doğru bir şekilde anlamamızı olanak tanıyan bir teknik olarak karşımıza çıkar (Agresti, 1984). Çalışmada bu teknikle cinsiyet yanlışlığına bakılarak alanyazına katkı sağlayacağı düşünülmektedir.

Başarı testi öğrencilerin öğrenme düzeylerini ölçme ve değerlendirme açısından önemli olduğu kadar öğretmenlere de öğrencilerin konuya ne kadar ögrendiklerini ve öğrenmenin hangi düzeyde olduğunu belirlemeye yardımcı olur (Türkdoğan vd., 2011). Bu da öğrencilere daha etkili bir şekilde rehberlik etmeyi sağlar. Vücutumuzdaki Sistemler ünitesi karmaşık, soyut ve çok yönlü bir konu olması sebebiyle öğrencilerin bu sistemleri anlamaları için testlerde yer alan soruların öğrenmelerini derinleştirmelerini sağlamaları açısından önemli olduğu düşünülmektedir. Ayrıca bu ünitede yer alan 5 adet Sistemler konularının kazanım sayıları oldukça yüksektir (11 kazanım) (Tablo 7). Bu da ünitenin 6. Sınıf öğrencilerin akademik gelişimi açısından oldukça önemli olduğu sonucunu doğurmuştur. Bu nedenle bu ünite seçilerek geçerli ve güvenilir bir başarı testi oluşturulmasına karar verilmiştir.

#### **Araştırmmanın amacı**

Araştırmada altıncı sınıf fen bilgisi dersindeki “Vücutumuzdaki Sistemler” ünitesine yönelik bir başarı testi geliştirmek amaçlanmıştır. Bu amaçla “Vücutumuzdaki Sistemler” ünitesi için geliştirilmiş çoktan seçmeli soruların geçerlik-güvenirlilik analizi yapılarak toplam 26 sorudan

oluşan 4 seçenekli bir başarı testi geliştirilmiş ve toplamda 157, 6.sınıf öğrencisine uygulanmıştır. Teste ilişkin istatistikî analizler yöntem bölümünde detaylı olarak sunulmuştur.

### **Yöntem**

Bu araştırmada altıncı sınıf Fen Bilimleri öğretim programındaki Vücutumuzdaki Sistemler ünitesiyle ilgili bir başarı testi geliştirilmiştir. Geliştirilen başarı testine ait güvenirlik ve geçerlik çalışmaları yapılmıştır. Başarı testiyle ölçülmesi hedeflenen ünite kazanımları (Tablo 7) Milli Eğitim Bakanlığı'nın (MEB) Fen Bilimleri dersi öğretim programından alınmıştır. Geliştirilen başarı testinde toplamda 26 adet soruya yer verilmiştir. Yöntem Akış Şeması (Şekil 1) aşağıda belirtilmiştir.

**Şekil 1**

*Yöntem akış şeması*



### **Çalışma grubu**

Bu araştırmada Güneydoğu Anadolu Bölgesindeki bir ilin iki merkez ilçesine bağlı biri Bilim ve Sanat Merkezi diğer devlet ortaokulu olmak üzere iki kurumda 2022-2023 akademik yılında öğrenim gören toplam 157 6. sınıf öğrencisi yer almıştır. Çalışma grubuna dair bilgiler Tablo 1'de verilmiştir.

**Tablo 1**

*Çalışma grubunun ilçelere ve cinsiyetlere göre dağılımı*

İlçeler	Erkek	Kız	Toplam
A	28	42	70
B	45	42	87
Toplam	73	84	157

**1. Test geliştirme aşamaları**

**a) Testin amacının belirlenmesi**

Test geliştirme sürecinde ilk olarak testin amacı belirlenerek geliştirilen başarı testi ile altıncı sınıf öğrencilerinin Fen Bilimleri dersi “Vücutumuzdaki Sistemler” ünitesindeki eksiklikleri belirlenerek testin madde ve test istatistiklerini hesaplamak amaçlanmıştır. Bu amaç doğrultusunda ilk aşamada ölçülmek istenen kazanımlar Fen Bilimleri dersi öğretim programından faydalılarak tespit edilmiştir. Geliştirilen başarı testinde öğrencilerin Fen Bilimleri becerilerini ölçen kazanımlarla ilgili maddeler geliştirilmiştir.

**b) Konunun belirlenmesi**

Günlük hayatı sıkılıkla karşılaştığımız vücutumuzun sistemler yapısına ilişkin konu ve kavramlar, bu çalışmada geliştirilen başarı testine dahil edilmiştir. Bu ünitede öğrenciler; Destek ve Hareket Sistemi, Sindirim Sistemi, Dolaşım Sistemi, Solunum Sistemi ve Boşaltım Sistemi ile ilgili organları ayırt edecektir.

**c) Madde havuzu**

6.sınıf "Vücutumuzdaki Sistemler" ünitesine ait konulara ilişkin öğrencinin bilimsel süreç becerileri ve kazanımları dikkate alınarak, sınav soruları araştırmacılar tarafından araştırmanın amacına uygun olarak geliştirilmiştir. Fen Bilimleri ders kitapları, soru çözüm bankaları ve bilgisayar destekli ortamlardan yararlanılarak ilk olarak 50 maddeden oluşan soru havuzu ile süreçte başlanmıştır.

**d) Yazım denetimi ve uzman görüşü alma**

50 maddenin yazım kurallarını ve bilimsel doğruluğunu değerlendirmek için üç Fen Bilimleri öğretmeni, bir Biyoloji ve bir Türkçe öğretmeni ile Eğitimde Ölçme ve Değerlendirme alanında uzmanın görüşleri alınmıştır. Ayrıca soruların doğruluğu ve belirli bir kazanıma uygunluğu konusunda da görüş alışverişinde bulunulmuştur. Kazanım hedefleri, öğrenci düzeyi ve konunun bütünlüğünü kapsayıp kapsamadığı, yazım kurallarına ve öğrenci seviyesine uygunluğu incelenmiş ve uzman görüşleri doğrultusunda soru sayısı 26'ya indirilmiştir (Tablo 2). Yaklaşık 11 kazanım bulunduran ünitede her kazanıma yaklaşık iki adet soru düşmektedir. Yalnızca Destek ve Hareket Sistemi ve Sindirim Sistemi'ne ait birer kazanıma dört adet soru düşmektedir.

Her soru 3,8 puandır, doğru yanılışı götürmemektedir ve testin cevaplanma süresi bir ders saatı ile (40 dk) sınırlanmıştır. Araştırmada başarı testi hazırlanırken iki farklı MEB onaylı 6. sınıf fen bilimleri ders kitabı ve 6. sınıf EBA etkinlikleri ve sanal Fen Bilimleri siteleri kullanılmıştır. Uzmanlar, bir sorunun tek kelime sebebiyle iki doğru cevabının olabilmesi, bazı sorularda belirsiz, anlaşılmayan ifadelere yer verilmesi ve bir soruda şekilde gösterilen organ yerinin net olmaması şeklinde geri dönütler vermişlerdir. Madde köklerinde ve seçeneklerde düzenlemeler yapılması gereği de ayrıca önerilmiştir. Uzman görüşleri dikkate alınarak sorular düzenlenmiş ve testteki soru sayısı 26'ya indirilmiştir. Uzman görüşlerinin alınmasından sonra test, sekiz 6. Sınıf öğrencisine uygulanmış ve testin 40 dakikalık süre içerisinde cevaplanıldığı görülmüştür. Öğrencilerin büyük bir çoğunluğu uygulanan teste ilişkin olumsuz bir görüşü şeiksel kaynaklı olarak bildirmiştir.

**Tablo 2**

*Uzmanlara ait bilgiler*

Alanı	Unvanı	Kıdemİ
Eğitimde Ölçme ve Değerlendirme	Doç. Dr.	16 yıl
Fen Bilimleri	Öğretmen	10 yıl
Biyoloji	Öğretmen	21 yıl
Türkçe	Öğretmen	11 yıl

**e) Testin kapsam geçerliğinin incelenmesi**

Başarı testinin kapsam geçerliliği için; uygulamanın öncesinde Fen Bilimleri Programının (MEB, 2018) "Vücutumuzdaki Sistemler" ünitesinin bölümleri ve kazanımları belirtke tablosunda gösterilmiştir (Tablo 5).

**f) Testin uygulanması**

Geliştirilen başarı testi iki farklı devlet ortaokuluna ulaştırılmış ve toplamda 157 öğrenciye yüz yüze uygulanmıştır. Testi doğru cevaplayan öğrencilerin cevapları Excel sayfasına doğru yanıtlar "1", yanlış ve boş bırakılan yanıtlar ise "0" şeklinde yazılmıştır. Böylelikle bir katılımcının testten alabileceği en yüksek puan 26 doğru ile (100), en az puan ise sıfır olarak belirlenmiştir.

**g) Madde güçlük ve ayırt edicilik analizi**

**Madde güçlüğü (pj)**

Bir sınav sorusuna doğru cevap veren kişi sayısının sınava giren toplam kişiye oranına madde güçlüğü denir. Her soru için doğru cevap oranını gösterir.

*p'nin değer aralıkları;*

-Madde güçlüğü 0-1 arasında olup 1'e yaklaştıkça madde kolaylaşır, 0'a yaklaştıkça madde zorlaşır.

-Bir testte kolay, orta ve zor güçükte madde olmalıdır.

-Madde güçüğünün 0.50 seviyesinde olması istenilen durumdur (Büyüköztürk 2021).

### **Madde ayırt ediciliği ( $r_{jx}$ )**

Ayırt edicilik, test görevinin ve çeldiricilerin bilen bir öğrenci ile bilmeyen bir öğrenciyi ayırt edebildiği seviyedir. Yani bir maddenin başarı düzeyi en fazla olan grubu en zayıf gruptan ayırt edip etmediğini belirlemek için kullanılan ölçütür. Test sorularının seçiciliği yüksek olmalıdır. Ayırt ediciliği çok az olan maddeler testten çıkarılmalıdır.

Bir maddenin ayırt ediciliği de  $-1 < r_{jx} < 1$  arasında değişir (Ergül, 2022).

#### **h) Güvenirlilik analizi**

Güvenilirliği hesaplama yöntemleri değişkendir ve KR-20 ve KR-21 istatistikleri, doğru yanıtın “1” puan, yanlış yanıtın “0” puan aldığı testler için oluşturulmuştur. KR-20, her bir delilin zorluğu bilindiğinde, KR-21 ise bilinmediğinde kullanılır. Öğrenci test puanlarını doğrulamak için kullanılabilecek başka bir istatistiksel teknik ise Cronbach's Alpha yöntemidir. Bu yöntem KR-20 ile aynı mantıga dayandığı için doğru yanıt için bir puan, yanlış yanıt için sıfır puan verilen testlerde kullanılabilmektedir (Atılgan, 2013).

#### **Eşdeğer yarılar analizi**

Bu yöntemde, test bir grup katılımcıya uygulanır ve testi puanlamadan önce test maddeleri iki eşdeğer forma bölünür. Her durumda konum sayısı aynı olmalı ve şekil eşdeğer olmalıdır (Crocker ve Algina, 1986).

Anastasi ve Urbina (1997), eşdeğer yarı güvenirliğin hesaplanmasındaki çözüm adımı, testi ikiye bölgerek iki eşdeğer form oluşturmak olduğunu belirtmiştir. Crocker ve Algina (1986), testi iki parçaya ayırmak için “4” yöntem önermiştir. Bunlar; kota eşitlik yöntemi, madde zorluğuna dayalı kota eşitlik yöntemi, rastgele atama yöntemi ve içerik karşılaştırma yöntemi. Bu çalışmada rastgele seçilen maddelerden birinin eşdeğer formlardan birincisine, diğerinin ikinciye yerleştirilebildiği rastgele atama yöntemi performans testi geliştirilmiştir. Rastgele bir atamada, her ögenin eşdeğer formlara yerleştirilme olasılığı eşit ve bağımsızdır. Magnusson (1968), benzer uygulamalı bir testte, bu yöntem kullanılarak iki eşdeğer şeklin yapılabileceğini bulmuştur. İçerik eşleştirme yöntemi, incelenmekte olan davranışlara dayalı olarak bir testteki öğeleri eşleştirir. Eşlerden biri birinci, diğeri ikinci biçimde sınıflandırılır. Magnusson (1968), test maddelerinin zorluğu göz önüne alındığında, testin heterojen olması durumunda bu yöntemi kullanmanın daha doğru olacağını savunmuştur.

Bir testi herhangi bir ikiye bölgeme yöntemiyle böldükten sonra, eşdeğer yarımların güvenilirliğini tahmin etmek için Spearman-Brown düzeltme formülü, Rulon formülü, Guttman formülü ve Flanagan formülü kullanılabilir. Testi iki parçaya ayırip iki yarı oluşturuktan sonra, her bir yanıtlayıcının iki yarıdaki performansı arasındaki korelasyon Pearson'ın Çarpım Momentleri katsayısı kullanılarak hesaplanır ve bu katsayı yarı test güvenirlik katsayısıdır (Crocker & Algina, 1986). Tüm test için güven katsayısı, Spearman-Brown düzeltme formülü kullanılarak hesaplanır. Spearman-Brown düzeltme formülü aşağıdaki eşitsizlikteki gibidir.

$$rtt = 2r_{12} / 1 + r_{12}$$

“ $r_{tt}$  = Testin tümüne dair güvenirlik katsayısı”

“ $r_{12}$  = Yarı teste dair güvenirlik katsayısı”

### ***Yanlılık analizleri:***

Yanlılık test maddelerin testi cevaplayan bir grup öğrenci bakımından avantaj sağlaması olarak tanımlanmaktadır. Böyle bir durumda testin geçerliği etkilenecektir ve dolayısıyla testin her bir maddesinin yanlışlık analizlerinin yapılması ve raporlanması istenecektir. Geliştirilen Fen Bilimleri başarı testinin cinsiyete göre yanlışlık analizleri yapılmıştır. Klasik test kuramına göre yapılan yanlışlık analizlerinde Mantel-Haenszel Tekniği (MH) kullanılmış ve test maddelerinin cinsiyet değişkenine göre farklılaşan madde fonksiyonu (DMF) gösterip göstermedikleri belirlenmiştir.

#### ***Mantel- Haenszel teknigi (MH):***

Bu teknikte odak grup olarak kız öğrenciler (toplam 83 öğrenci), referans grup olarak ise erkek öğrenciler (toplam 74 öğrenci) alınmıştır. MH teknliğinde odds (aMH) büyülüğünün değerine göre maddenin hangi grubun lehine davranışının belirlenir. Eğer;

aMH =1 olduğu zaman madde DMF içermez.

aMH >1 olduğunda madde referans gruba avantaj sağlamaktadır.

aMH <1 olduğunda ise madde odak gruba avantaj sağlamaktadır.

MH teknliğinde delta istatistiğinin ( $\Delta MH$ ) değerine göre DMF'nin etki büyülüğü belirlenir.  
Eğer;

$|\Delta MH| < 1$  ise DMF ihmali edilebilir (A Düzeyi)düzyedendir. $1 \leq |\Delta MH| < 1,5$  ise DMF orta (B Düzeyi) düzeydedir .

$|\Delta MH| \geq 1,5$  DMF yüksek (C Düzeyi) düzeydedir ( Gök, Kelecioğlu & Doğan, 2010).

#### ***i) Maddelerin sınıflanması***

Her bir test ögesinin bilimsel süreç boyutuna ve bilgi boyutuna göre sınıflandırma tablosu Yenilenmiş Bloom Taksonomisine göre oluşturulmuştur (Tablo 6-7).

Verilerin analizi için Optik okuyucu (Testokur), JASP 0.15, FACTOR 11.05, TAP (Test Analysis Program) 19.1.4 ve Optik Okuyucu (Testokur) programları kullanılmıştır. Analiz sonuçları bulgular kısmında verilmiş ve yorumlanmıştır.

#### ***Bulgular***

Geliştirilen başarı testinin madde ve test istatistikleri hesaplanmıştır. Öncelikle madde ve test istatistikleri detaylı olarak incelenmiş daha sonra ise güvenilirlik ve geçerlik kanıtları ile test maddelerin yenilenmiş Bloom taksonomisine göre durumları sırasıyla sunulmuştur.

#### ***Madde ve test istatistikleri***

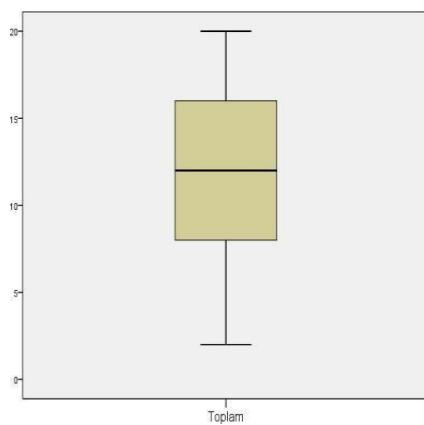
Öncelikle maddelere ait betimsel istatistikler hesaplanmıştır. Başarı testinin betimsel istatistikleri (Tablo 3) ve kutu grafiği (Şekil 2) ve Histogram Grafiği (Şekil 3) incelendiğinde testin normal dağılım gösterdiği (çarpıklık = ,424; basıklık = -0,522) (Tabachnick & Fidell, 2007) ve üç değerler içermediği görülmektedir.

**Tablo 3**  
*Betimsel istatistikler*

<b>Betimsel İstatistikler</b>	<b>İstatistik Değerleri</b>
<b>Ortalama</b>	$12,809 = 49,3\%$
<b>%95 Güvenirlik Aralığında Ortalama</b>	<b>Alt Sınır:</b> 0,820 <b>Üst Sınır:</b> 0,885
<b>Medyan</b>	$12,00 = 46,2\%$
<b>Varyans</b>	32,21
<b>Standart sapma</b>	5,675
<b>Minimum</b>	$1,00 = 3,8\%$
<b>Maksimum</b>	$26,00 = 100\%$
<b>Çarpıklık</b>	0,424
<b>Basıklık</b>	-0,522

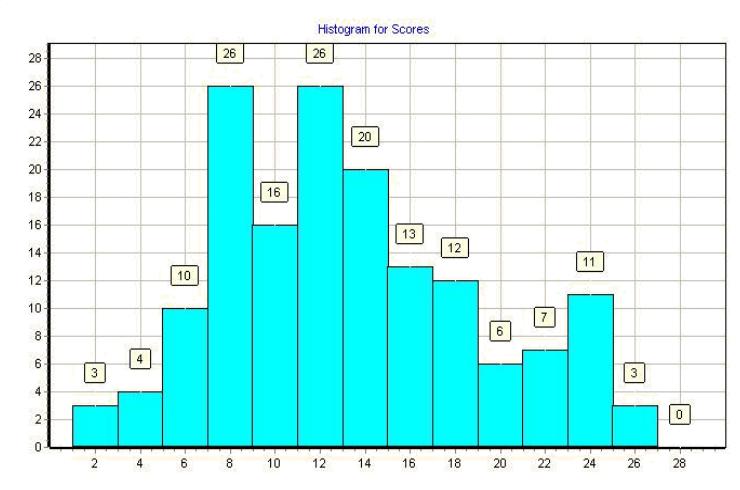
Tablo 3'e göre ortalama 12,809, medyan 12, varyans 32,21, standart sapma 5,675, çarpıklık 0,424, basıklık değeri -0,522'dir.

**Şekil 2**  
*Kutu grafiği*



**Şekil 3**

*Başarı testinin histogram grafiği*



Maddelere ait güçlük ve ayırt edicilik değerleri TAP programı aracılığıyla hesaplanmıştır. Madde güçlük- ayırt edicilik değerleri Tablo 4'te verilmiştir.

**Tablo 4**

*Madde güçlük ve ayırt edicilik değerleri*

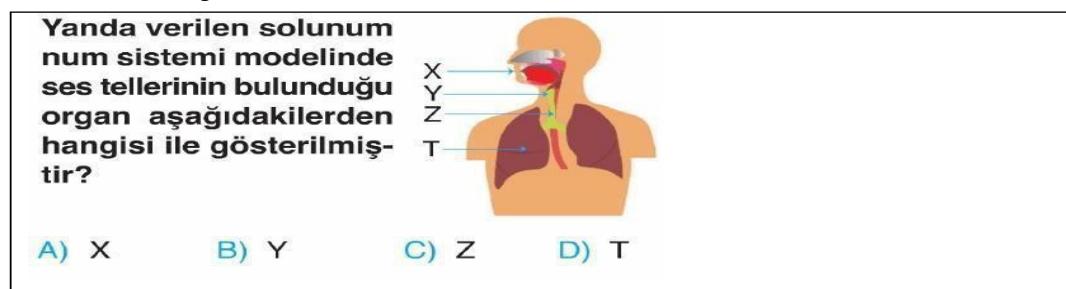
Madde No	Madde Güçlüğü	Madde Ayırt Ediciliği	Madde No	Madde Güçlüğü	Madde Ayırt Ediciliği
$m_1$	0,35	0,27	$m_{14}$	0,69	0,79
$m_2$	0,55	0,64	$m_{15}$	0,52	0,53
$m_3$	0,55	0,64	$m_{16}$	0,30	0,38
$m_4$	0,52	0,35	$m_{17}$	0,65	0,72
$m_5$	0,75	0,44	$m_{18}$	0,35	0,66
$m_6$	0,33	0,53	$m_{19}$	0,30	0,32
$m_7$	0,36	0,18	$m_{20}$	0,71	0,50
$m_8$	0,61	0,59	$m_{21}$	0,15	-0,3
$m_9$	0,46	0,78	$m_{22}$	0,50	0,66
$m_{10}$	0,46	0,37	$m_{23}$	0,68	0,54
$m_{11}$	0,60	0,82	$m_{24}$	0,49	0,44
$m_{12}$	0,48	0,69	$m_{25}$	0,62	0,72
$m_{13}$	0,38	0,53	$m_{26}$	0,45	0,69

Maddelere ait güçlük değerlerinin 0,15 ile 0,75 arasında olduğu görülmüştür. Testin ortalama güçlük değeri 0,50 olarak hesaplanmıştır. Buna göre orta güçlüğe bir test geliştirildiği söylenebilir. Test maddelerinin ayırt edicilik değerlerinin ise -0,03 ile 0,79 değerleri arasında değiştiği görülmüştür. Testin ortalama ayırt edicilik indeksi 0,527 olarak hesaplanmıştır. Ayırt edicilik indeksi 0,19 ve altında olan maddelerin testten çıkarılarak 0,20 ile 0,29 arasında olan maddelerin ise gerekli düzeltmelerinin yapılip teste alınması önerilmektedir (Hasançebi vd., 2020). Bu durumda test maddelerinden sadece 21. maddenin ayırt ediciliğinin önerilen değerin altında olduğu, 7. maddenin ise sınır değerde olduğu görülmektedir. Dolayısıyla 21 numaralı madde testten çıkarılmıştır.

Test maddeleri analizleri incelendikten sonra, çoğu çeldiricinin istenen düzey aralığında olduğu, üst gruplar ve alt gruplar arasında istenilen boyutta yaptığı görülmüştür. Şekil 4'teki 21. madde için çeldirici, grubu istenen düzeyde ayırt edemediği görülmüştür. Aşağıda testteki 21. madde numarasına göre veriler açıklanmıştır. 21. maddenin ayırt edicilik tablosu Tablo 5'te gösterilmiştir.

**Şekil 4**

21. Soru görseli



**Tablo 5**

21. Maddenin alt ve üst grupta madde analizi

MADDE	Seçenek	A	B	C*	D
21	<b>Toplam</b>	22(0,140)	102 (0,650)	23*(0,146)	10 (0,064)
	<b>En Yüksek</b>	2 (0,043)	38 (0,826)	6 (0,130)	0 (0,000)
	<b>En Düşük</b>	10(0 , 233)	20,(0,465)	7 (0,163)	6 (0,140)
	<b>Güçlük</b>	-8 (-0 ,189)	18*(0,361)	-1(-0,163)	-6(-0,140)

Tablo 5'e göre 21.maddenin doğru cevabı C seçeneği olmasına rağmen şekilde Y ve Z ile gösterilen kısımda soluk borusu ile gırtlağın yeri net olmadığı için üst gruptan sadece 6 kişi bu soruya doğru cevap vermiş olup üst gruptan 38 kişi ise çeldirici B'yi işaretlemiştir. Bu nedenle bu maddede şekilde gösterilen Y ve Z ile belirtilen organların kısımları daha net gösterilecektir.

### **Testin güvenirligine ilişkin bulgular**

Maddelerin güclüklerinin birbirinden farklı olması ve maddelerin ikili puanlanması (1-0) sebebiyle iç tutarlılık katsayıları olarak KR-20 değerine bakılmış ve KR-20 değeri 0,84 olarak bulunmuştur. İç tutarlılık katsayıları 1'e yakın olduğu ölçüde test güvenilir olduğundan (Büyüköztürk, 2021) geliştirilen testin güvenirlüğünün yüksek olduğu söylenebilir ( $KR20 > 0,7$ ).

#### **Eşdeğer yarılar:**

Eşdeğer yarılar analizi yapılrken ilk olarak JASP\_ ta korelasyondan Spearman Korelasyon değeribulunarak (Tablo 6) formülle testin tümüne ilişkin güvenirlilik katsayısı hesaplanmıştır.

**Tablo 6**

Spearman korelasyon

Değişken	Değer
Eşdeğer yarılar Kişi sayısı	157
Spearman's katsayısı	0,703
Anlamlılık değeri	< .001

$$rtt = 2r_{12} / 1 + r_{12}$$

$r_{tt}$  = Testin tümüne ilişkin güvenirlilik katsayısı

$r_{12}$  = Yarı teste ilişkin güvenirlilik katsayısı

$$r_{tt} = 2 \times 0,703 / 1 + 0,703 = 0,83$$

1'e yaklaşığı için 0,83 değeri yüksek güvenirliktedir (Büyüköztürk, 2021).

### **Testin geçerliğine ilişkin bulgular**

Bir ölçme aracının amaca hizmet etme derecesi “geçerlik” olarak tanımlanır (Erkuş vd., 2017). Bir ölçme aracının geçerliğinin sağlanması için geçerliğin alt boyutlarının ayrı ayrı analiz edilmesi ve geçerlik kanıtlarının sunulması gereklidir.

#### **a. Kapsam Geçerliği**

Kapsam geçerliğinin sağlanması için test geliştirme sürecinde belirtke tablosunun oluşturulması ve geliştirilen soruların bu kazanımlara göre hazırlanması gereklidir. Hazırlanan soruların uzmanlara sunulması ve uzman görüşlerinin alınması ile testin kapsam geçerliği sağlanır. Geliştirilen Fen Bilimleri başarı testine ait belirtke tablosuna bakıldığından kazanımlara uygun dağılımlı sorular ile temsil edildiği görülmekte ve kapsam geçerliği sağlanmaktadır. Tablo 7'de ise geliştirilen başarı testine ait belirtke tablosu verilmiştir. Fen Bilimleri Dersi Öğretim Programı (2018) baz alınarak 6. Sınıf ikinci ünite olan Vücutumuzdaki Sistemler ünitesine ait belirtke tablosu aşağıda verilmiştir (Tablo 7).

**Tablo 7.**

*6. Sınıf vücutumuzdaki sistemler ünitesine ait belirtke tablosu*

Konu No	Kazanımlar	Soru
<b>Destek ve Hareket Sistemi</b>	6. 2. 1. 1. “Destek ve hareket sistemi ile ilgili yapıları örneklerle gösterir.” 6. 2. 2. 1. “Sindirim sistemini oluşturan yapı-organların görevlerini modelleriyle açıklar.”	1-2-4-5 3-10
<b>Sindirim Sistemi</b>	6. 2. 2. 2. “Besinlerin kana geçebilmesi için fiziksel-kimyasal sindirim uğraması gerektiğini çıkarır.” 6. 2. 2. 3. “Sindirimde yardımcı organların mekanizmasını açıklar.”	6-7-9-11 16-8-12
<b>Dolaşım Sistemi</b>	6. 2. 3. 1. “Dolaşım sistemi ile ilgili yapı-organların görevlerini modelle açıklar.” 6. 2. 3. 2. “Büyük- küçük kan dolaşımını şema üzerinde inceler ve görevini açıklar.” 6. 2. 3. 3. “Kanın yapısını-görevlerini açıklar.” 6. 2. 3. 4. “Kan grupları arasındaki kan alışverişini gösterir.” 6. 2. 3. 5. “Kan bağışının toplum açısından önemini değerlendirdir.”	13-25 16-19 15-18 14-17 20-24
<b>Solunum Sistemi</b>	6. 2. 4. 1. “Solunum sistemini ile ilgili yapı-organların görevlerini modeller ile açıklar.”	21-25
<b>Boşaltım Sistemi</b>	6. 2. 5. 1. “Boşaltım sistemiyle ilgili yapı-organları model üzerinde görevlerini özetler.”	22-26

Tablo 7’de yer alan kazanımlar Fen Bilimleri Öğretim Programında “Vücutumuzdaki Sistemler” ünitesine ait olup Fen Bilimleri Öğretim Programından (2018) alınmıştır.

**b. Yapı geçerliği**

Yapı geçerliliği ölçme açısından maddelerin birbirleriyle yüksek ilişkisinin olduğu, maddelerin benzer özellikleri ölçüfü koşullarda sağlanır. Yapı geçerliği kanıtlarının elde edilmesinde Açımlayıcı Faktör analizden yararlanılır (Erkuş vd., 2017). Faktör analizi işlemlerinden

önce maddelere ait ayırt edicilik indeksleri hesaplanmış, düşük ya da ters ayırt edicilik değerlerine sahip maddeler belirlenmiştir. Ayırt ediciliği düşük olan 2 1. Madde faktör analizine dahil edilmemiştir. Analiz FACTOR 11.5 programı kullanılarak gerçekleştirilmiştir. Faktör analizine ilişkin sonuçlar Tablo 8 ve 9'da verilmiştir.

**Tablo 8**

*“KMO ve Bartlett testi” sonuçları*

“KMO Örneklem Yeterliliği” Testi		0,848
Yaklaşık Ki Kare		
Değeri		1690.5
Bartlett'in Küresellik Testi		
sd		300
p		,00001

Tablo 8'deki analiz sonuçlarına göre geliştirilen başarı testinin yeterli örneklem büyülüğüne sahip olduğu söylenebilir ( $KMO>0,60$ ) (Büyüköztürk, 2021). Bartlett küresellik testi anlamlıdır ( $p<,05$ ). Analiz sonuçlarına göre faktör analizinin temel koşulları sağlanmıştır.

**KMO ölçütü**

KMO'nun özel bir formatı olan aynı zamanda her bir madde için elde edilen örneklem uygunluk ölçüsü Tablo 8'de görüldüğü gibi 0.848'dir. Bu durum yeterli örneklem büyülüğüne ulaşlığı sonucuna ulaşılabilir.

**Tablo 9.**

*Faktör analizi sonuçları*

Yapı	Özdeğer	Açıklanan Varyans (%)	Yığmalı Varans(%)
1	8.76369	0.35055	0.35055
2	2.21853	0.08874	0.43929
3	1.74332	0.06973	
4	1.5035	0.06014	
5	1.33774	0.05351	
6	1.12081	0.04483	
7	1.09719	0.04389	
8	0.7353	0.03873	
9	0.9682	0.03688	
10	0.87992	0.03520	
11	0.75193	0.03008	
12	0.67782	0.02711	

13	0.50245	0.0201
14	0.4585	0.01834
15	0.40709	0.01628
16	0.34372	0.01375
17	0.29132	0.01165
18	0.23538	0.00942
19	0.21015	0.00841
20	0.19122	0.00765
22	0.15966	0.00639
23	0.11171	0.00447
24	0.08306	0.00332
25	0.02096	0.00084
26	0.00002	0

Faktör programına 21. Sorunun ayırt ediciği – değerde olduğu için bu madde testten çıkartılarak faktör programına analize verilmiştir. Analiz sonuçlarına göre FACTOR programı yapı sayısını 1 olarak önermiştir. Dolayısıyla geliştirilen başarı testinin tek bir yapıyı ölçüdüğü görülmektedir. Tek faktör toplam varyansın %35’ini açıklamaktadır.

#### c. Yanlılık analizleri

Geliştirilen başarı testinde yanlışlık analizlerinden biri olan Mantel- Haenszel Tekniği (MH) kullanılmıştır. Cinsiyet, bir yanlışlık analizinde genellikle seçilen bir demografik özelliktir çünkü cinsiyet, bireyler arasında önemli bir ayırmayı sağlayabilir. Yanlılık analizinde cinsiyetin seçilmesi, verileri değerlendirdirken toplumda var olan cinsiyet eşitsizliklerini ve yanlışlıklar gözleme ve analiz etme fırsatı sağlar (Bakan Kalaycıoğlu ve Kelecioğlu, 2011). Toplumda cinsiyet rolleri ve beklenileri farklılık gösterebilir. Bu nedenle, cinsiyetin seçimi bir yanlışlık analizinde dikkatli bir şekilde ele alınmalıdır. Vücutumuzdaki Sistemler ünitesi ile ilgili yapılan geçmiş başarı testi çalışmalarında MH teknigi ile cinsiyet yanlışlığı çalışmalarına rastlanmadığından çalışmada kıyaslama durumu yapılamamıştır.

Bu teknikte odak grup olarak kız öğrenciler (toplam 83 öğrenci), referans grup olarak ise erkek öğrenciler (toplam 74 öğrenci) alınmıştır. MH teknigi ile (aMH) büyülüğünün değerine göre maddenin hangi grubun lehine davranışının belirlenmiştir.

Geliştirilen Fen Bilimleri başarı testine ait yanlışlık analizi için referans aralıkları yöntem kısmında belirtilmiş olup ilgili sonuçlara Tablo 10'da deñinlmıştır.

**Tablo 10**

*Cinsiyete göre yanlışlık analizi sonuçları*

Madde No	aMH Değeri	Avantajlı Grup	ΔMH  Değeri	DMF Düzeyi
Madde <sub>1</sub>	0,9014	Kız	0,106	İhmal

Madde <sub>2</sub>	0,9993	Kız	0,0007	edilebilir(A) İhmal
Madde <sub>3</sub>	0,7697	Kız	0,2672	edilebilir(A) İhmal
Madde <sub>4</sub>	0,8397	Kız	0,1783	edilebilir(A) İhmal
Madde <sub>5</sub>	0,4237	Kız	0,8764	edilebilir(A) İhmal
Madde <sub>6</sub>	0,9354	Kız	0,0682	edilebilir(A) İhmal
Madde <sub>7</sub>	1,8807	Erkek	-0,6446	edilebilir(A) İhmal
Madde <sub>8</sub>	0,5672	Kız	0,5787	edilebilir(A) İhmal
Madde <sub>9</sub>	0,8835	Kız	0,1265	edilebilir(A) İhmal
Madde <sub>10</sub>	1,9801	Erkek	-0,6972	edilebilir(A) İhmal
Madde <sub>11</sub>	0,4561	Kız	0,8011	edilebilir(A) İhmal
Madde <sub>12</sub>	1,1842	Erkek	-0,1726	edilebilir(A) İhmal
Madde <sub>13</sub>	0,8633	Kız	0,15	edilebilir(A) İhmal
Madde <sub>14</sub>	0,4607	Kız	0,791	edilebilir(A) İhmal
Madde <sub>15</sub>	1,2727	Erkek	-0,2461	edilebilir(A) İhmal
Madde <sub>16</sub>	1,2522	Erkek	-0,2295	edilebilir(A) İhmal
Madde <sub>17</sub>	0,5339	Kız	0,6405	edilebilir(A) İhmal
Madde <sub>18</sub>	1,7723	Erkek	-0,5841	edilebilir(A) İhmal
Madde <sub>19</sub>	1,696	Erkek	-0,5392	Edilebilir(A)

Madde <sub>20</sub>	0,4297	Kız	0,862	İhmal Edilebilir(A)
Madde <sub>21</sub>	29,951	Erkek	-11,196	İhmal Edilebilir(A)
Madde <sub>22</sub>	0,8348	Kız	0,1843	İhmal Edilebilir(A)
Madde <sub>23</sub>	0,4968	Kız	0,714	İhmal Edilebilir(A)
Madde <sub>24</sub>	0,9302	Kız	0,0738	İhmal Edilebilir(A)
Madde <sub>25</sub>	0,7449	Kız	0,3005	İhmal Edilebilir(A)
Madde <sub>26</sub>	12,308	Erkek	-0,2119	İhmal Edilebilir(A)

Analiz sonuçlarına göre test maddelerin cinsiyete göre DMF düzeylerinin ihmali edilebilir seviyede (A düzeyi) olduğu görülmüştür.

**d. Yenilenmiş Bloom Taksonomisine göre test maddelerinin sınıflandırılması**

Bümen'in (2010) çalışmasında da bahsedildiği gibi yenilenen taksonomide, Bloom taksonomisindeki bilgi, kavrama, analiz-sentez basamakları yenilenerek sırasıyla hatırlama, anlama, çözümleme-yaratma olarak tekrar isimlendirilmiştir. Buarada alt basamaklar vurgulanarak program öğeleri arasındaki tutarlılık sağlanmış ve son olarak örnek ölçme işlemlerine yer verilerek taksonomi son halini almıştır. Yenilenmiş Bloom taksonomisinin bu avantajları göz önünde bulundurularak bu çalışmada geliştirilen başarı testi yenilenmiş Bloom taksonomisine uygun yapılandırılmıştır. İki ayrı fen bilimleri öğretmeni, bir Türkçe öğretmeni ve bir ölçme ve değerlendirme uzmanı her maddeleri birbirinden bağımsız bir şekilde maddeleri kategorize etmiştir. Daha sonra bir araya gelerek maddeleri hangi sınıflamaya kim, neden koyduğuna dair açıklamalar yaparak bilgi alışverişinde bulunmuşlardır. Son olarak maddelerin bilişsel süreç ve bilgi boyutunu ortak bir tabloda sınıflandırılmışlardır. Başarı testindeki çoktan seçmeli soruların Yenilenmiş Bloom Taksonomisi'ne göre yerleşimleri Tablo 11'de gösterilmiştir.

**Tablo 11**

*“Yenilenmiş Bloom Taksonomisine” Göre Testte Yer Alan Maddelerin Bilişsel Süreç ve Bilgi Boyutu*

Bilişsel Süreç Boyutu						
Bilgi Boyutu	Hatırlama	Anlama	Uygulama	Çözümleme	Değerlendirme	Yaratma

1. Olgusal Bilgi	1, 4, 5, 9, 12, 13, 15, 21, 22, 23
2. Kavramsal Bilgi	10,24,25
3. İşlemsel Bilgi	3, 6      2,8,16,18 19,26
4. Üstbilişsel Bilgi	7,11,14,17,20

#### Tartışma ve Sonuç

Bu çalışmanın amacı, 6. sınıf Vücutumuzdaki Sistemler ünitesi ile ilgili geçerli ve güvenilir bir test geliştirmektir. Bu amaçla MEB Fen Bilimleri Dersinin öğretim programındaki kazanımları incelenmiş ve test ile ölçümek istenen hedefler belirlenmiştir. Kazanımlar belirlendikten sonra test sorularının görselleri Paint programından faydalananlarak hazırlanmıştır. Test uygulaması sonuçlarına göre testin KR20 değeri 0,84, ortalama güclüğü 0,493 ve ortalama ayırt edicilik değeri 0,527 olarak kaydedilmiştir. Kaydedilen sonuçlara göre orta güclükte bir başarı testi geliştirildiği söylenebilir.

26 maddelik teste 157 öğrenci girmiştir. En düşük yapan öğrencinin 1 doğru cevabı olup 3,8 puan almıştır. En yüksek yapan öğrencinin 26 doğru cevabı olup 100 puan almıştır. Medyan 12 olup puan karşılığı 100 üzerinde 46,2dir. Aritmetik Ortalama 12.8 olup 46,3 puana karşılık gelmektedir. Standart sapması 5,6 Varyansı 32,2. Çarpıklık-Basılıklık katsayısı 0,424-0,522 yani -1 ile +1 arasındadır. Normal dağılım göstermektedir. 26 madde TAP programında analiz edilmiştir. Testin ortalama güclüğü 0,493 olup orta güclükte bir testtir. Ayırt ediciliklerin ortalaması 0,527 (0,4 ün üzerinde olduğu için iyi maddelerdir). Nokta çift serili korelasyon değeri 0,453 KR-20 güvenirlik iç tutarlılık katsayısı 0,848 olduğu için güvenilir bir test geliştirildiği söylenebilir. KR21 0,83; Spearman Korelasyon Değeri; 0,709 olup eş değer yarılar testin güvenirlik değeri 0,83'tür. Testin güvenirlüğinin yüksek olduğu söylenebilir. Üst grup en başarılıdan sıralanmaya başlandığında 46. kişiden kesilmiş olup en az 16 doğrusu vardır. Alt grup ise en başarısızdan sıralanmaya başlandığında 43. kişiden kesilmiş olup en yüksek 8 doğrusu bulunmaktadır. Çıkarılması gereken madde 1 tane'dir (21.Madde). 21. Madde Factor programına dahil edilmemiştir. . Analiz sonuçlarına göre FACTOR programı yapı sayısını 1 olarak önermiştir. Dolayısıyla geliştirilen başarı testinin tek bir yapıyı ölübü gibi görülmektedir. Tek faktör toplam varyansın %35'ini açıklamaktadır.

Başarı testinin geliştirilmesi aşamasında, Şekil 1'de gösterilen akış şeması takip edilmiştir. Alanyazına bakıldığından çoğu başarı geliştirme çalışmalarında çalışmamıza benzer test geliştirme aşamalarının izlendiği görülmüştür (Haladyna, 1997; Kargin ve Gül, 2021; Keçeci vd., 2019; Açıkgöz ve Karslı, 2015; Bolat ve Karamustafaoglu, 2019; Kızkapan ve Bekaş, 2018). Bu durumun çalışmamıza paralellik sağlayarak başarı testi uygulama sürecinin güvenirliğini artırdığı düşünülmektedir.

Alanyazına bakıldığından, Vücutumuzdaki Sistemler ünitesi kapsamında hazırlanmış çeşitli 178

başarı testi araştırmalarının çalışmamızda paralellik gösterdiği görülmektedir. Bu çalışmalar:

Kargin ve Gü'lün (2021) yaptığı “Vücutumuzdaki Sistemler ve Sağlığı” ile ilgili geliştirdiği başarı testi geliştirme çalışmalarında 46 madde, 212 yedinci sınıf öğrencisine uygulanmıştır. Madde analizinden elde edilen sonuca göre çoktan seçmeli sorudan oluşan (40 tane) “testin ortalama güclüğü 0,61” ve “ortalama ayırt ediciliği ise 0,47” olarak hesaplanmış olup sonrasında bazı sorularda düzeltme yapılmıştır ve güvenilirlik analizi için farklı çalışma grubuna (178 kişi) uygulanan başarı testinin “KR-20 değeri 0,86” olarak hesaplanmıştır.

Keçeci, Yıldırım ve Zengin (2021) çalışmalarında aynı ünite ile ilgili başarı testinde 36 maddeden oluşan testi 7 uzman görüşüne sunduktan sonra 30 'a düşürülmüştür. 158, 7.sınıf öğrencisine pilot uygulama yapılmış olup madde analizi yapılmış ve 0,19 ve düşük 7 madde çıkarılmıştır.23 madde ile tekrar pilot uygulama 100 öğrenciyle yapılmış “Croanbach Alfa katsayısı 0,83, Spearman Brown iki yarı test korelasyon değeri 0,81” olarak hesaplanan geçerli ve güvenilir başarı testi hazırlanmıştır. Çetinkaya ve Taş (2016), 3 aşamalı “Vücutumuzda Sistemler” ünitesine yönelik kavram yanılışlarını belirleyen geçerli-güvenilir olarak geliştirdikleri başarı testini 357 öğrenciye uygulamıştır. Toplam 17 maddelik ve 51 sorudan oluşan kavram tanı testinde önce “KR-20 değeri 0,774” hesaplanmış daha sonra üç aşama toplam değerlendirildiğinde ise” güvenilliği 0,683” olarak bulunmuştur.

Alanyazında Fen Bilimleri dersi için geliştirilen başarı testleri incelendiğinde test geliştirme aşamalarında benzer aşamalarının gerçekleştirildiği görülmektedir (Kargin ve Gü'l (2021), Keçeci, Yıldırım ve Zengin (2021), Çetinkaya ve Taş (2016)). Test geliştirme sürecinde öncelikle kazanımların belirlenmesi daha sonra test sorularının hazırlanması ve son olarak madde ve test istatistiklerinin sunumu bu aşamaları oluşturmaktadır. Genel olarak çalışmalarda başarı testi geliştirilirken soru yazımı aşamasında Bloom'un taksonomik süreçlerinin yeterince rapor edilmediği, sadece Çetinkaya ve Taş (2016) 'nın tabloya döktükleri görülmüştür. Alanyazınınındaki bu eksiklikleri gidermek amacıyla bu araştırmada başarı testinin soruları Yenilenmiş Bloom Taksonomisi'ne uygun olarak geliştirilmiştir.

Alanyazın incelendiğinde diğer eksik görülen unsurlardan birisi ise Fen Bilimleri dersi için geliştirilen başarı testlerinde yanlışlık analizlerinin yapılmamasıdır. Yanlılık, testi oluşturan maddelerin bir grup öğrenci bakımından avantaj oluşturması olarak tanımlanabilir. Böyle bir durum testin geçerliğini doğrudan etkileyecektir. Test amaca hizmet etmeyecek ve test sonuçlarına hatalar karışacaktır. Gerçekleştirilen bu araştırmada bir başarı testi geliştirilirken yanlışlık analizlerinin nasıl yapılabileceği detaylı olarak sunulmuştur. Cinsiyete dayalı yanlışlık analizleri Klasik Tepki Kuramı'na dayalı Mantel Haentzel tekniği kullanılarak gerçekleştirilmiş ve maddelerin düşük düzey (A) DMF gösterdiği görülmüştür. Bu da ihmäl edilebilir boyuttadır. Bu durum testin geçerliğine ilişkin önemli bir kanıt sunmaktadır. Alanyazında Fen Bilimleri dersiin geliştirilen başarı testlerinin Kargin ve Gü'l, (2021), Keçeci vd., (2021), Çetinkaya ve Taş'ın (2016) yaptığı çalışmalarda yanlışlık analizlerinin yapılmadığı görülmektedir. Bu durum bu testlerinin geçerliği için bir tehdit oluşturmaktadır.

Bu çalışmanın sonuçlarına göre ölçme aracının, alanyazında yer alan “6. sınıf Fen Bilimleri dersi” “Vücutumuzdaki Sistemler” ünitesine ait sistem, sonuç ve kavramlara katkı sağlayabilecek geçerli ve güvenilir başarı testi olduğu söylenebilir. Geçerlik ve güvenilirlik analizleri yapılmış

ölçme araçlarının eğitimde kullanılması önemlidir (Gömleksiz ve Erkan, 2010). Ölçme işlemi sonucunda öğrencilerin doğru bir şekilde değerlendirebilmesi ölçümede kullanılan aracın istenilen amaca uygun şekilde hizmet etmesine bağlıdır (Ergül, 2022). Geliştirilecek başarı testlerine ait farklı geçerlik kanıtlarının sunulması bu nedenle çok önemlidir. Sonuç olarak geliştirilen bu başarı testiyle Fen Bilimleri eğitiminde kullanılabilecek geçerli ve güvenilir bir ölçme aracının geliştirilme süreci detaylı olarak betimlenmiştir. Bu çalışmaya ek olarak bu sayede farklı illerde ve okullarda benzer çalışmalar yapılabilir; başarı testinin evreni temsil etme yeteneği geliştirilebilir. Araştırmacılara çeşitli yanılık analizi teknikleri kullanarak geliştirdikleri başarı testleri için kanıt sağlayabilir. Çalışma, daha fazla madde havuzu oluşturarak tekrarlanabilir. Bu, maddelerin kapsam geçerliğini artırmaya olanak tanır. Çalışmamızda düzeltilmesi gereken bir madde (21.madde) testten çıkarılmıştır. Bu maddeler geliştirilebilir ve kullanılır hale getirilebilir. Ayrıca, test sonuçlarına bağlı olarak, öğretmenler öğrencilerin eksik ve yanlış bilgileri düzeltmeleri için zenginleştirilmiş öğrenme ortamları ve teknikleri geliştirebilirler.

#### **Araştırmacıların Katkısı**

Bu çalışmada; B. AKAR, bilimsel yayın araştırması, veri analizi, makalenin oluşturulması; İ. GÜNEŞ çalışmanın uygulanması ve veri toplama İ. YILDIRIM, çalışma tasarımları, veri analizi ve makalenin genel kontrolünün yapılarak koordine çalışmalarında katkı sağlamışlardır.

#### **Çıkar Çatışması**

Yazarlar tarafından herhangi bir çıkar çatışması beyan edilmemiştir.

#### **Kaynakça**

- Açıkgöz, M., & Karslı, F. (2015). Alternatif ölçme-değerlendirme yaklaşımları kullanılarak iş ve enerji konusunda geliştirilen başarı testinin geçerlilik ve güvenirlilik analizi. Amasya Üniversitesi Eğitim Fakültesi Dergisi, 4(1), 1-25 <https://doi.org/10.17539/aej.71851>
- Agresti, A. (1984). Analysis of ordinal categorical data. New York: John Wiley & Sons Anastasi, A., & Urbina, S. (1997). *Psychological testing (7th ed.)*. New York: Macmillan Pub.Co.Inc.
- Aybek, E. C., Yaşar, M., & Kartal, S. (2021). Öğretmen Yapımı Bir Testteki Maddelerin Değişen Madde Fonksiyonu Bağlamında İncelenmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, (52), 281-300. <https://doi.org/10.9779/pauefd.825631>
- Bahar, M., Nartgün, Z., Durmuş, S., & Bıçak, B. (2012). *Geleneksel tamamlayıcı ölçme değerlendirme teknikleri* (5. Baskı). Ankara: Pegem Akademi.
- Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology, Statistical Section*, 3, 77-85. *Bilim Dergisi*, Cilt 35, Sayı 156.
- Bekdemir, M., Selim, Y. (2008). Revize Edilmiş Bloom Taksonomisi ve Cebir Öğrenme Alanı Örneğinde Uygulaması. *Erzincan Eğitim Fakültesi Dergisi*, 10(2), 185-196. <https://dergipark.org.tr/en/download/article-file/67433>
- Bümen, N. T. (2010). Program geliştirmede bir dönüm noktası: Yenilenmiş Bloom taksonomisi. *Eğitim ve Bilim*, 31(142). <http://213.14.10.181/index.php/EB/article/view/837/189>
- Bolat, A., & Karamustafaoglu, S. (2019). Vücutumuzdaki sistemler ünitesi başarı testi geliştirme: Geçerlik ve güvenirlilik. *Gazi Eğitim Bilimleri Dergisi*, 5(2), 131-159.

<https://dergipark.org.tr/en/download/article-file/773031>

Büyüköztürk, Ş. (2021). *Sosyal bilimler için veri analizi el kitabı* (29. Baskı). Ankara: Pegem Akademi.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt Rinehart and Winston.

Çetinkaya, M., & Erol, T. A. Ş. (2016). Vücutumuzda Sistemler Ünitesine Yönerek Üç Aşamalı Kavram Tanı Testi Geliştirilmesi. *Ordu Üniversitesi Sosyal Bilimler Enstitüsü Sosyal Bilimler Araştırmaları Dergisi*, 6(15), 317-330. <https://dergipark.org.tr/en/download/article-file/273724>

Dinçer L., Yıldırım İ. (2022), Yenilenmiş Bloom Taksonomisine Göre Başarı Testi Geliştirme Çalışması: 5. Sınıf “Hello” Ünitesi Örneği. *International Journal of Language Academy*.

Erkuş, A., Sünbül, Ö., Sünbül, S., Yormaz, S. & Aşiret, S. (2017). *Psikolojide Ölçme ve Ölçek Geliştirme-II* (2 bs.). Ankara: Pegem Akademi Yayıncılık

Gök, B., Kelecioğlu, H. & Doğan, N. (2010). Değişen madde fonksiyonunu belirlemeye Mantel-Haenszel ve lojistik regresyon tekniklerinin karşılaştırılması, *Eğitim ve Bilim Dergisi*, Cilt 35, Sayı 156 <http://eb.ted.org.tr/index.php/EB/article/view/19>

Gömlekşiz, M., & Erkan, S. (2010). *Eğitimde ölçme ve değerlendirme* (2. Baskı). Ankara:Nobel Yayın Dağıtım.

Gönen, S., Kocakaya, S., & Kocakaya, F. (2011). Dinamik konusunda geçerliliği ve güvenilirliği sağlanmış bir başarı testi geliştirme çalışması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 8(1), 40-57. <https://dergipark.org.tr/en/download/article-file/146252>

Haladyna, T. M. (1997). Writing test items to evaluate higher order thinking. London: Allyn & Bacon.

Hasançebi, B., Terzi, Y., & Küçük, Z. (2020). Madde güçlük indeksi ve madde ayırt edicilik indeksine dayalı çeldirici analizi. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, 10(1), 224-240.

Ergül, D. (2022, Eylül). *Eğitimde ve Psikolojide Ölçme ve Değerlendirme*. Ankara Üniversitesi. <https://acikders.ankara.edu.tr/course/view.php?id=9575>

Karasar, N. (2012). *Bilimsel araştırma yöntemleri* (24. baskı). Ankara: Nobel Yayınevi

Kargin, P. D., Gül, Ş. (2021). Development of an achievement test for sixth grade body systems and health unit, *Ihlara Journal of Educational Research*, 6(1), 1–26. <https://dx.doi.org/10.47479/ihed.729412>

Keçeci, G., Yıldırım, P., & Kirbağ Zengin, F. (2019). Sistemler akademik başarı testi: Geçerlik ve güvenirlilik çalışması. *Ulusal Eğitim Akademisi Dergisi (UEAD)*, 3(1), 96-114. <https://doi.org/10.32960/uead.514188>

Kızkapan, O., & Bektaş, O. (2018). Fen eğitiminde başarı testi geliştirilmesi: Hücrebolünmesi ve kalıtım örneği. *Uluslararası Eğitim Bilimleri Dergisi*, 2(1), 1-18. <https://dergipark.org.tr/en/download/article-file/499076>

- Magnusson, D. (1968). Test theory. Massachusetts. Addison-Wesley.
- MEB (2018), Fen Bilimleri Dersi Öğretim Programı (İlkokul ve Ortaokul 4,5,6,7 ve 8. Sınıflar), Ankara.
- Ogan Bekiroğlu, F. (2004). *Ne kadar Başarılı? Klasik ve Alternatif Ölçme- Değerlendirme Yöntemleri ve Fizikte Uygulamalar* (1. baskı). Ankara: Nobel Yayın Dağıtım.
- Tabachnick, F. & Fidell, L. (2007) *Using Multivariate Statistics* (Fifth Edition). California State University/Northridge. Pearson Education.
- Türkdoğan, A. , Şahin, S. M. & Baki, A. (2011). Süreç Değerlendirmesinde Elde Edilen Kavram Yanılgılarının Test Geliştirme Çalışmasında Kullanılması . Elektronik Sosyal Bilimler Dergisi, 10 (37) , 78-92 . <https://dergipark.org.tr/tr/pub/esoder/issue/6151/82615>
- Yıldırım, İ., & Demir, S. (2016). Oyunlaştırma temelli Öğretim İlke ve Yöntemleri dersi öğretim programı hakkında öğrenci görüşleri. *Uluslararası Eğitim Programları ve Öğretim Çalışmaları Dergisi*, 2(6), 85-102.
- Yıldız, S., Keçeci, G., & Kirbağ-Zengin, F. (2019). Dengeli beslenme akademik başarı testi: geçerlik ve güvenilirlik araştırması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 16(1), 848-868. <https://dergipark.org.tr/en/download/article-file/887524>

#### EXTENDED ABSTRACT

##### Introduction

Science is a branch of science that explains the elements that connect the phenomena in the universe with scientific laws and helps to understand nature and the relationships in nature. Science curricula in our country have been developed from the behaviorist approach to the constructivist approach based on research and inquiry as of 2014 and lastly revised in 2018. According to the revised curriculum, constructivist learning theory has come to the fore (Gömleksiz, 2005). The degree of internalization of the goals, tasks and results determined by the student in science lessons, as well as their achievements, should be determined by valid and reliable measurement tools (Gönen, Kocakaya, & Kocakaya, 2011). There have been criticisms that Bloom's Taxonomy is in one dimension, that the synthesis level is more complex than the evaluation, and that Synthesis even includes Evaluation. A new taxonomy was needed to keep up with the developments in the world, to differentiate teaching methods and techniques, to take into account the psychology of development and learning, and to develop measurement and evaluation techniques (Bumen, 2010). Anderson et al. (2006) created a Revised Bloom's Taxonomy with six steps as Recall, Comprehension, Application, Analysis, Evaluation and Creation. It was aimed to develop an achievement test according to this taxonomy.

##### Purpose of the Study:

The aim of the study was to develop an achievement test for the "Systems in Our Body" unit in the sixth grade science course. For this purpose, a 4-option achievement test consisting of a total of 26 questions was developed by analyzing the validity and reliability of the multiple-

choice questions developed for the "Systems in Our Body" unit and applied to a total of 157 6th grade students. Statistical analysis of the test is presented in detail in the method section. It is thought that the developed achievement test will contribute to the literature.

#### **Method**

In this study, an achievement test was developed for the Systems in Our Body (Unit 2) unit in the sixth grade Science curriculum. Reliability and validity studies were conducted for the developed achievement test. The unit achievements targeted to be measured by the achievement test were taken from the Ministry of National Education's (MoNE) Science curriculum.

#### **Study Group**

In this study, a total of 157 6th grade students studying in two institutions, one of which is a Science and Art Center and the other is a public secondary school, in two central districts of a province in the Southeastern Anatolia Region of Turkey between the years 2022-2023 took part.

#### **Test development stages were as follows:**

First, the purpose of the test was determined. The subject was determined. Item pool was created. Spelling was checked and expert opinion was obtained. Content validity was examined in the specification table developed in accordance with the learning outcomes and expert opinion. The test was applied. Item difficulty and discrimination indices were examined. Reliability analysis was performed. Equivalent halves analysis and item side.