ATATURK
UNIVERSITY
PUBLICATIONS

Educational Academic Research

Formerly: Atatürk University Journal of Kazım Karabekir Education Faculty
Official journal of Atatürk University Faculty of Education

Issue 46 • September 2022

EISSN 2822-3535
education-ataunipress.org

Educational Academic Research

AMAÇ VE KAPSAM

Educational Academic Research bağımsız, tarafsız ve çift kör hakemlik ilkelerine uygun olarak yayınlanan bilimsel, açık erişimli bir süreli yayındır.

Educational Academic Research 2003 yılında yayın hayatına başlamıştır. Dergi, Atatürk Üniversitesi Kâzım Karabekir Eğitim Fakültesi'nin sadece internet üzerinden yayınlanan resmi yayınıdır ve Mart, Haziran, Eylül ve Aralık aylarında üç ayda bir yayınlanır. Dergi Türkçe ve İngilizce olarak yayınlanmaktadır.

Dergi, eğitim ve öğretimle ilgili başta meta-analiz ve meta-sentez çalışmaları olmak üzere nitel ve nicel araştırma makaleleri, derlemeler ve çağdaş eğitim literatürü ile ilgili kuramsal çalışmalar yayınlar. Derginin temel amacı, eğitim alanında üretilen bilimsel bilgileri geniş bir platformda paylaşmaktır. Bu doğrultuda dergi, araştırmacıları, eğitim uygulayıcılarını ve eğitim politikacılarını ortak bir noktada buluşturmayı amaçlamaktadır. Dergi, eğitim sorunlarının çözümüne, eğitimde eğitimin kalitesine ve yeni bilgilerin üretilmesine yol açacak araştırmalara odaklanmaktadır. Dergiye öğretmen adayları ile yapılan çalışmalar kabul edilmemektedir.

Derginin editöryel ve yayın süreçleri Council of Science Editors (CSE), Committee on Publication Ethics (COPE), European Association of Science Editors (EASE) ve National Information Standards Organization (NISO) kılavuzlarına uygun olarak biçimlendirilmiştir. Educational Academic Research' ün editöryel ve yayın süreçleri, Akademik Yayıncılıkta Şeffaflık ve En İyi Uygulama (doaj.org/bestpractice) ilkelerine uygun olarak yürütülmektedir.

Educational Academic Research, China National Knowledge Infrastructure (CNKI) ve TÜBİTAK ULAKBİM TR Dizin tarafından indekslenmektedir.

Derginin tüm masrafları Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi tarafından karşılanmaktadır. Makale değerlendirme ve yayın işlemleri için yazarlardan ücret talep edilmemektedir. Tüm makaleler https://education-ataunipress.org/ sayfasındaki online makale değerlendirme sistemi kullanılarak dergiye gönderilmelidir. Derginin yazım kurallarına, gerekli formlara ve dergiyle ilgili diğer bilgilere web sayfasından erişilebilir.

Dergide yayınlanan makalelerde ifade edilen bilgi, fikir ve görüşler Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi, Baş Editör, Editörler, Yayın Kurulu ve Yayıncı'nın değil, yazar(lar)ın bilgi ve görüşlerini yansıtır. Baş Editör, Editörler, Yayın Kurulu ve Yayıncı, bu gibi yazarlara ait bilgi ve görüşler için hiçbir sorumluluk ya da yükümlülük kabul etmemektedir.

Educational Academic Research yayınlanma modeli Budapeşte Açık Erişim Girişimi (BOAI) bildirgesine dayanan açık erişimli bilimsel bir dergidir. Derginin arşivine https://education-ataunipress.org/ adresinden ücretsiz olarak erişilebilir. Educational Academic Research'nin içeriği, Creative Commons Alıntı-GayriTicari 4.0 (CC BY-NC) Uluslararası Lisansı ile yayınlanmaktadır.

Yazarlara Bilgi'nin güncel versiyonuna https://education-ataunipress.org/ adresinden ulaşabilirsiniz.

Editör: Ali Osman ENGİN
Adres: Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi, Erzurum, Türkiye
Tel: +90 442 231 35 50
E-posta: aliosman.engin@atauni.edu.tr

Yayıncı: AVES
Adres: Büyükdere Cad., 105/9 34394 Şişli, İstanbul, Turkey
Telefon: +90 212 217 17 00
E-posta: info@avesyayincilik.com
Web: www.avesyayincilik.com

Educational
Academic
Research

CONTENTS / İÇİNDEKİLER

RESEARCH ARTICLES / ARAŞTIRMA MAKALELERİ

1 **Üniversite Öğrencilerinde Pozitiflik ve Psikolojik İyi Oluş Arasındaki İlişkide Umudun Aracı Rolü**
Mediating Role of Hope in the Relationship Between University Students' Positivity and Psychological Well-Being
Mehmet Buğra ÖZHAN, Nazife ÜZBE ATALAY

11 **"Bitki ve Hayvanlarda Üreme, Büyüme ve Gelişme" Konusuna Yönelik Bir Başarı Testinin Geliştirilmesi**
The Development of an Achievement Test Towards "Reproduction, Growth and Development in Plants and Animals" Subject
Fatih YAZICI, Şeyda GÜL, Mustafa SÖZBİLİR

20 **Araştırmacı Öğretmen Kimliğinin Gelişim Süreci**
The Development Process of the Researcher-Teacher Identity
Kevser YILDIRIM, Şirin KARADENİZ

36 **Fen Bilimleri Öğretmenlerinin Mesleki Doyum, Öz Yeterlilik, Öğretim Kaygısı ve Okul Kültürü Düzeylerinin Çeşitli Değişkenlere Göre İncelenmesi**
Examination of Science Teachers' Levels of Job Satisfaction, Self-Efficacy, Teaching Anxiety, and School Culture According to Various Variables
Ataman KARAÇÖP, Tufan İNALTEKİN

49 **Okul Öncesi Öğretmenlerinin Matematik Kaygılarının Matematik Öz Yeterliklerine Olan Etkisinin İncelenmesi**
Investigation of Preschool Teachers' Math Anxiety Effect on Mathematics Self Efficiency
Yunus AYTAÇ, Merve ÜNAL

62 **İlkokul Öğrencilerinin Fiziksel Aktivite Düzeylerinin ve Problem Çözmeye İlişkin Yansıtıcı Düşünme Becerileri**
Physical Activity Levels and Reflective Thinking Skills for Problem-Solving of Primary School Students
Elif AYDİN, Bijen FİLİZ, Mehmet DURNALİ

71 **Fen Bilgisi Öğretmen Adaylarının Astronomiye Yönelik Tutumlarının ve Astronomi Öğretimi Özyeterlik İnançlarının İncelenmesi: Çevrim İçi Astronomi Eğitimi**
Investigation of Pre-Service Science Teachers' Attitudes Toward Astronomy and Self-Efficacy Beliefs in Astronomy Teaching: Online Astronomy Education
Sedat KARAÇAM, Döndar YENER, Sedef CANBAZOĞLU BİLİCİ, Çiğdem ŞAHİN ÇAKİR, Nejla YÜRÜK

85 **Ortaöğretim Yetkinliklerini Karşılama Düzeyi Açısından 2009 Sosyoloji Ders Programının Değerlendirilmesi**
Evaluation of the 2009 Curriculum of Sociology Lesson in Terms of Meeting Secondary Education Competencies
Ümmühan Güliz ORHAN, Abdullah ADIGÜZEL

REVIEWS / DERLEMELER

99 **IQ: Nedir? Ne Değildir?**
IQ: What is it? What is it not?
Furkan ATMACA, Sema TAN

106 **Pedagojik Alan Bilgisi ve Teknolojik Pedagojik Alan Bilgisine İlişkin Yapılan Çalışmaların Meta-Sentez Yöntemiyle Analizi**
Analysis of Studies on Pedagogical Content Knowledge and Technological Pedagogical Content Knowledge by Meta-Synthesis Method
Hüseyin YOLCU, Dilan KAYA DURNA, Ali AKAN, Şafak ULUÇINAR SAĞIR

ATATÜRK ÜNİVERSİTESİ YAYINLARI
ATATURK UNIVERSITY PUBLICATIONS

Research Article *Araştırma Makalesi* DOI: 10.5152/AUJKKEF.2022.837043

"Bitki ve Hayvanlarda Üreme, Büyüme ve Gelişme" Konusuna Yönelik Bir Başarı Testinin Geliştirilmesi

The Development of an Achievement Test Towards "Reproduction, Growth and Development in Plants and Animals" Subject

Fatih YAZICI[1]
Şeyda GÜL[2]
Mustafa SÖZBİLİR[3]

[1]Millî Eğitim Bakanlığı, Erzurum, Türkiye
[2]Atatürk Üniversitesi, Kazım Karabekir Eğitim Fakültesi, Biyoloji Eğitimi Anabilim Dalı, Erzurum, Türkiye
[3]Atatürk Üniversitesi, Kazım Karabekir Eğitim Fakültesi, Kimya Eğitimi Anabilim Dalı, Erzurum, Türkiye

ÖZ

Bu araştırma, altıncı sınıf fen bilimleri dersinde yer alan 'Bitki ve Hayvanlarda Üreme, Büyüme ve Gelişme' konusuna yönelik bir başarı testi geliştirmek amacıyla yürütülmüştür. Bu nicel araştırmanın verileri tarama yöntemi ile toplanmıştır. Testin geliştirilmesi sürecinde öncelikle alan yazın taraması yapılarak fen bilimleri dersi öğretim programındaki kazanımlara uygun 30 adet çoktan seçmeli soru belirlenmiştir. Testin geçerlik çalışmalarında söz konusu sorular alanında uzman akademisyen ve öğretmenlerin görüşüne sunulmuştur. Uzman görüşleri sonrası sorular dil, kapsam, anlaşılabilirlik vb. açılardan incelenmiş, gerekli görülen durumlarda yazım ve içerik açısından düzeltmeler yapılmış, ayrıca yaş grubu da dikkate alınarak soru sayısı 20'ye indirilmiştir. Hazırlanan 20 soruluk taslak test Erzurum il merkezindeki bir ortaokulun altıncı sınıfında öğrenim gören 58 öğrenciye uygulanmıştır. Sonrasında değerlendirmeye alınan toplam 55 öğrenciden (29 kız, 26 erkek) elde edilen verilere madde analizi yapılmıştır. Madde analizi sonucunda testin ortalama güçlüğü 0,66 ve ortalama ayırt ediciliği ise 0,56 olarak hesaplanmıştır. Ayrıca testte kalması gerektiği düşünülen bazı sorularda düzeltmeler yapılmıştır. Madde analizi sonrası testin güvenirlik çalışmaları ise 60 kişilik farklı bir öğrenci grubu ile yapılmıştır. Güvenirlik analizi sonucunda başarı testinin KR-20 güvenirlik katsayısı 0,93 olarak hesaplanmıştır. Geçerlik ve güvenirlik analizleri sonucunda son şekli verilen nihai testin Bloom Taksonomisi'ne göre farklı zorluk seviyelerinde sorular içeren, geçerli ve güvenilir bir test olduğu görülmüştür.

Anahtar Kelimeler: Başarı testi, üreme, test geliştirme

ABSTRACT

This research was implemented to develop an achievement test toward "Reproduction, Growth and Development in Plants and Animals" subject. The data of this quantitative research were collected through survey design methodology. In the test development process, the literature was firstly reviewed and then 30 multiple choice questions in accordance with learning outcomes in the science course curriculum was determined. For the validity study, the questions in test were presented to the opinion of academicians and teachers who are experts in their fields. After expert opinions, questions were examined in terms of language, scope, understandability, etc., and corrections were made in terms of spelling and content when necessary. Finally, the number of questions was reduced to 20 by taking the age group into account. The draft test with 20 questions was applied to 58 students who attended the sixth grade of a secondary school in the city center of Erzurum. Subsequently, item analysis was performed on the data obtained from 55 students (29 girls, 26 boys) who were evaluated. In the item analysis, the average difficulty of the test was calculated as 0.66 and the average discriminant was 0.56. In addition, some questions that were thought to remain in the test were corrected. After the item analysis, the reliability studies of the test was performed with a different group of 60 students. According to the reliability analysis, the KR-20 reliability coefficient for achievement test was 0.93. As a result of the validity and reliability analysis, it was seen that the final test was a valid and reliable test that includes questions at different difficulty levels of Bloom's Taxonomy.

Keywords: Achievement test, reproduction, test development

Geliş Tarihi/Received: 07.12.2020
Kabul Tarihi/Accepted: 28.04.2022

Sorumlu Yazar/Corresponding author: Fatih YAZICI
E-mail: fatyaz20@gmail.com

Cite this article as: Yazıcı, F., Gül, Ş., & Sözbilir, M. (2022). The development of an achievement test towards "Reproduction, Growth and Development in Plants and Animals" subject. *Educational Academic Research*, 46, 11-19.

Content of this journal is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Giriş

Günlük hayatla ilişkilendirme ve öğrenme kavramlarını bir arada inceleyen çalışmaların son yıllarda yoğunluk kazandığı görülmektedir. Elbette bu durum şaşırtıcı değildir. Zira söz konusu olan bu iki kavramına bakıldığında, aralarında yakın bir ilişki olduğu ve öğrenmenin gerçekleşmesinde günlük yaşamın etkili bir faktör olduğu açıkça görülmektedir (Ürey & Cerrah Özsevgeç, 2015).

Günlük yaşantımızda sıkça karşılaştığımız sıradan olayların pek çoğunun aslında günlük yaşamımızda fen bilimlerinin bir yansıması olarak karşımıza çıkmaktadır. Bununla beraber yaşamımızla böylesine iç içe olan fen bilimleri alanına öğrencilerin yeterli düzeyde değer vermediği görülmektedir (İlhan & Hoşgören, 2015; Tekbıyık & Akdeniz, 2010). Bu durumu destekler nitelikte, alan yazında yürütülen pek çok çalışmada fen bilimlerinin günlük hayattan uzak, formüller ve genel kurallar yığını şeklinde algılandığı, ayrıca bilimsel gerçeklikten uzak bir biçimde ezberlenmek suretiyle geçiştirilmeye çalışıldığı ifade edilmektedir (Gül, 2019; İlhan & Hoşgören 2015; Şahin ve ark., 2013). Elbette günümüzde öğrencilere verilen fen eğitiminin temel amacı onlara mevcut bilgileri sunmak değil, aksine onlara bilgiye ulaşmanın temel yollarını öğretmeye çalışmaktır. Bu sayede yaparak ve yaşayarak bilgileri öğrenen bireyler bilimsel süreç becerilerini geliştirmenin yanında günlük yaşamlarında karşılaşabilecekleri yeni durumlara ve problemlere de etkili çözümler üretebilirler. Bu süreçte aktarılan bilginin ne kadarının öğrenciler tarafından anlaşıldığının belirlenmesi ise ancak başarı düzeylerinin ölçülmesi ile mümkün olabilmektedir (Bingöl & Halisdemir, 2017; Duruk ve ark., 2017; Güngörmez & Akgün, 2018).

Eğitimde başarı denildiğinde çoğu zaman akademik başarı kavramı kastedilmektedir. Akademik başarı, öğretim sürecinde öğrencilere kazandırılması hedeflenen beceri veya bilgilerin öğretmenler tarafından notlarla, test puanlarıyla ya da ikisi birlikte kullanılarak ifade edilmesi olarak tanımlanır (Akt. Akandere ve ark., 2010). Öte yandan öğrenme ortamında öğrencilerin akademik başarılarının düzeyi ise ölçme değerlendirmenin yapılmasıyla saptanabilir (Küçükahmet, 2004). Bu noktada özellikle ölçmenin gündelik yaşantımızda olduğu gibi eğitimde de önemli bir kavram olduğu karşımıza çıkmaktadır (Demir ve ark., 2016). Bireyler, ölçme sürecini kullanarak birçok karar alabilirler. Bu sebepten ölçme, insan faktörünü de amaçları arasına alan eğitim sürecinin önemli bir parçasını oluşturmaktadır. Eğitim süreci okul çatısı altında ele alındığında ise her bir öğrencinin belirli bir niteliğe sahip olup olmama durumu veya bu niteliğe sahip olma derecesi ölçme süreci ile tespit edilmeye çalışılır (Atılgan, 2013). Öğretmenler bu süreçte öncelikli olarak dersleri ile ilgili belirli hedefler belirleyerek bu hedeflere ulaşabilmek amacıyla uygun öğretim durumları oluştururlar. Belirlenen hedeflere ulaşma düzeyini ise yapmış oldukları ölçme sonuçlarını kullanarak saptarlar. Elbette bu saptamanın sağlıklı olabilmesi için belirli niteliklere nitelikli ölçme araçlarına ihtiyaç vardır. Alan yazın incelendiğinde ölçme aracı olarak farklı tipte testlerin kullanıldığı görülmektedir (Çepni ve ark., 2009). Örneğin öğrencilerin başarılarının bağımlı değişken olarak incelendiği çalışmalarda çoğunlukla başarı testleri kullanılmaktadır (Çardak & Selvi, 2018). Bununla beraber başarı testlerinin de farklı şekillerde hazırlandığı bilinmektedir. Elbette farklı şekillerde hazırlanan bu ölçme araçlarının her biri, kullanma amacına göre üstün ya da eksik olan farklı yönlere sahip olabilir. Öğrencilerin başarı düzeyleri ölçülürken belirli bir ünite ya da birkaç ünitenin yer aldığı konuların öğretildiği derslerin tamamını belirli bir sürede ve fazla sayıda soru sorularak etkili bir şekilde değerlendirmede çoktan seçmeli testler oldukça elverişlidir. Bu nedenle de çoktan seçmeli testler, günümüzde birçok eğitim kurumunda sıklıkla kullanılan ölçme araçları arasında sayılabilir (Akbulut & Çepni, 2013). Öğretmenlerin sınıftaki uygulamalarından ulusal düzeyde yapılan sınavlara kadar farklı düzey ve öğretim alanlarında çoktan seçmeli testlerin kullanıldığına sıklıkla rastlanmaktadır. Bu nedenle çoktan seçmeli testler özellikle iyi bilinmesi gereken bir ölçme aracıdır (Atılgan ve ark., 2011).

Öğrencilerin edinilen bilgileri hatalı veya eksik öğrenmelerinin saptanmasında kullanılan çoktan seçmeli testler gerek zaman gerekse hazırlanma ve geliştirilme açısından diğer birçok ölçme araçlarına kıyasla çok daha kullanışlı olmanın yanı sıra öğrenme ortamında da kolaylıkla uygulanabilmektedir. Puanlama ve uygulanabilme açısından diğer ölçme araçlarına göre daha kullanışlı ve kolay olan çoktan seçeli testler, çeldiricilerin de iyi hazırlanması halinde öğrencilerin eksik ve hatalı bilgilerinin tespiti ve sonuçlarının genellenebilmesini sağlamaktadır (Ayvacı & Durmuş, 2016; Demirci & Efe, 2007). Sıralanan bu avantajlarının yanında diğer testlerle karşılaştırıldığında, doğru cevabın bilinmeden şans eseri seçilebilmesi ise çoktan seçmeli testlerin dezavantajlarından biri olarak karşımıza çıkmaktadır (Mintzes ve ark., 2001). Bu dezavantajı en aza indirmek amacıyla iki aşamalı testler geliştirilmiştir. İki aşamalı testlerin çoktan seçmeli testlerden farklı olan yönü bu testlerin ikinci aşamasıdır. Bilindiği üzere ikinci aşama, testin ilk aşamasında verilen cevabın gerekçesi için bir açıklama istemektedir Böylece konu ile ilgili öğrencilerin ön bilgileri, konuyu anlama düzeyleri hatta kavram yanılgıları belirlenebilmektedir (Ayvacı & Durmuş, 2016). Özellikle kavram yanılgılarını belirlemede daha fazla kullanılması nedeniyle bu testlerin hazırlanmasının daha zor olduğu ve uzmanlık gerektirdiği söylenebilir. Bu nedenle bir aşamalı çoktan seçmeli olarak hazırlanan başarı testlerinin uygun analiz ve yöntemlerle geçerlik ve güvenirliğinin sağlanarak daha nitelikli hazırlanması son derece önemlidir. Bu noktada başarı testlerinin geliştirilmesinde standart bir başarı testinin geliştirilme aşamalarının izlenmesinde fayda vardır. Özellikle eğitim araştırmalarında, geliştirilen başarı testi ile öğrencilerin öğrenme düzeyleri, başarıları ile ilgili yorumlarda bulunulacak ve özellikle eğitim bilimlerine katkı anlamında önemli sonuçlara ulaşılabilecektir. Bu nedenle, başarı testi puanlarının kullanıldığı bilimsel çalışmalarda ancak geçerli ve güvenilir olduğu kanıtlanmış bir test ile bu tür çıkarımlarda bulunmak mümkün olmaktadır (Çardak & Selvi, 2018). Bu nedenle birçok alanda olduğu gibi fen eğitimi alanında da güvenirliği ve geçerliği sağlanmış testlerin oluşturulması son derece önemlidir (Özcan ve ark., 2019).

Fen eğitimi alanında başarı testi geliştirmeye yönelik birçok çalışma mevcuttur (Ayvacı & Durmuş, 2016; Demir ve ark., 2016; Güngörmez & Akgün, 2018; Özcan ve ark., 2019; Saraç, 2018). Bununla beraber fen bilimlerinin önemli bir alanı olan biyoloji konularına yönelik başarı testi geliştirme çalışmalarına bakıldığında ise daha sınırlı sayıda olduğu dikkati çekmektedir (Kaya & Gül, 2020; Kızkapan & Bektaş, 2018). Öte yandan fen bilimleri dersi kapsamında ele alınan biyoloji alanı ile ilgili önemli konulardan biri Bitki ve Hayvanlarda Üreme, Büyüme ve Gelişme (BHÜBG) konusudur. Bu konu 2013 yılı Fen Bilimleri Dersi Öğretim Programı (FBDÖP)'nda altıncı sınıf düzeyinde Canlılar ve Hayat konu alanına ait bir ünite olarak dört kazanım ve 12 ders saati şeklinde verilmekteydi (MEB, 2013). FBDÖP'nın MEB tarafından 2018 yılında güncellenmesiyle birlikte bu ünite yedinci sınıf Canlılar ve Yaşam konu alanı, Canlılarda Üreme, Büyüme ve Gelişme ünitesi adı altında Bitki ve Hayvanlarda Üreme, Büyüme ve Gelişme konusu olarak yerini almıştır. Bununla beraber bu

konuya yeni programda yedinci sınıf düzeyinde 12 ders saati olarak yer verilmesine rağmen kazanımlar aynı sayı ve içeriktedir. Bu noktada bu araştırmada geliştirilen testin uygulamaları her ne kadar 2018 yılı öncesinde altıncı sınıf düzeyindeki öğrenciler için yapılmış olsa da yeni sorulan içerik açısından programla uyumlu ve yedinci sınıf öğrencileri için de uygulanabilir olduğunu söylemek mümkündür. Burada asıl üzerinde durulması gereken nokta, alan yazın incelendiğinde hangi sınıf seviyesinde olursa olsun öğrencilerin bu konuya ait kavramları öğrenmede zaman zaman güçlükler yaşadığıdır (Arslan, 2007; Haras, 2009; Önder, 2011; Yip, 1998). Bu noktada öğrencilerin BHÜBG konusunun etkili öğretimine ek olarak yapılacak değerlendirme sürecinin de güvenilir bir şekilde yürütülebilmesinde geçerliği ve güvenirliği sınanmış ölçme araçlarına ihtiyaç vardır. Her ne kadar alan yazında Güneş ve Serdaroğlu (2020) tarafından bu üniteye ait bir başarı testi geliştirilmiş olsa da ilgili araştırmanın da 2013 yılı öğretim programına göre yapıldığı görülmektedir. Diğer taraftan bu araştırmada geliştirilen başarı testinin soru sayısının fazla olması, testin ortalama madde güçlüğünün kolay olması, soruların öğretim programındaki kazanımlara göre dağılımı hakkında bilgi verilmemesi, testteki soruların yenilenmiş Bloom Taksonomisine göre bilişsel alanın hangi basamaklarında yer aldığının incelenmemesi vb. nedenlerle bu üniteye yönelik farklı bir başarı testi geliştirilmesi uygun görülmektedir.

Araştırmanın Amacı

Bu araştırmanın amacı altıncı sınıfta öğrenim gören öğrencilerin fen bilimleri dersi 'Bitki ve Hayvanlarda Üreme, Büyüme ve Gelişme (BHÜBG)' konusundaki başarı düzeylerini ölçmeye yönelik geçerliği ve güvenilirliği sınanmış bir başarı testi geliştirmektir. Bu amaç doğrultusunda araştırmada, aşağıdaki araştırma sorularına cevap aranmıştır:

1. BHÜBG konusuna yönelik geliştirilen test geçerli midir?
2. BHÜBG konusuna yönelik geliştirilen test güvenilir midir?

Yöntemler

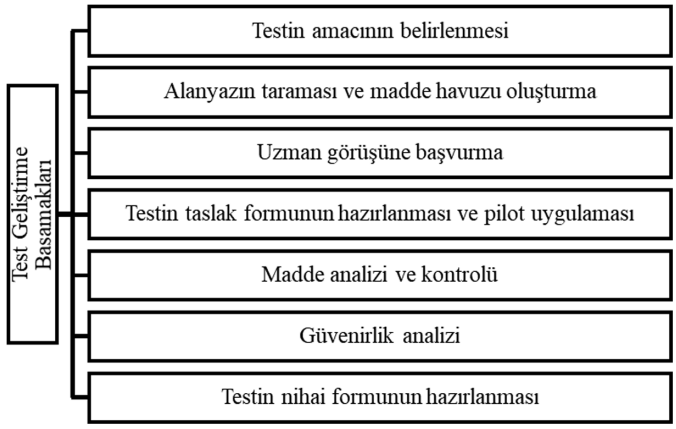
Araştırmanın bu bölümünde araştırma deseni, örneklem grubu, veri toplama süreci ve verilerin analizine yer verilmiştir.

Araştırma Deseni ve Örneklem

Nicel araştırma yaklaşımı temelinde kurgulanan bu araştırmada tarama yöntemi kullanılmıştır. Tarama araştırmaları insanların tutumları, inançları, değerleri, demografik özellikleri, davranışları, fikirleri, alışkanlıkları, arzuları vb. hakkında bilgi edinmek için kullanılır (McMillan & Schumacher, 2010). Bu araştırmada ise öğrencilerin başarı düzeylerini belirleyen bir test geliştirmek amacıyla tarama yöntemi tercih edilmiştir. Araştırmanın uygulamaları için Atatürk Üniversitesi Sosyal ve Beşeri Bilimler Etik Kurulu 'nun 04.09.2014 gün ve 48553601-5616 sayılı etik kurul izni alınmıştır. Araştırmanın örneklem grubu Erzurum İl merkezindeki kolay ulaşılabilir örnekleme yoluyla seçilmiş bir ortaokulda öğrenim gören ve daha önce bu konuyu işlemiş olan 58 altıncı sınıf öğrencisi olarak belirlenmiştir. Uygulamalar sonrasında 55 öğrenciden (29 kız, 26 erkek) elde edilen veriler değerlendirmeye alınarak madde analizi yapılmıştır. Madde analizi sonrasında ise testin güvenirlik analizi için farklı bir ortaokuldaki 60 kişilik öğrenci grubundan veri toplanarak analizler yapılmış ve teste nihai şekli verilmiştir.

Bitki ve Hayvanlarda Üreme, Büyüme ve Gelişme Başarı Testi (BHÜBGBT)'nin Geliştirilme Süreci

Alan yazın incelendiğinde, başarı testi geliştirilme sürecinde birtakım test geliştirme aşamalarının dikkate alınması gerektiği ileri sürülmektedir. Bu konu ile ilgili olarak Haladyna (1997) ve Kızkapan ve Bekaş (2018), bu süreçte dikkate alınması gereken aşamaları aşağıdaki gibi sıralamaktadır (Şekil 1). Bu araştırmada da BHÜBGBT'nin geliştirilme sürecinde Şekil 1'de gösterilen aşamalar izlenmiştir.



Şekil 1.
Test Geliştirme Aşamaları

Şekil 1 incelendiğinde test geliştirme sürecinin ilk aşamasının testin amacının belirlenmesi olduğu görülmektedir. Bu doğrultuda testin amacı, altıncı sınıf düzeyindeki öğrenciler için BHÜBG konusuna yönelik başarı testi geliştirmek olarak belirlenmiştir. İkinci aşamada alan yazın incelenerek ortaokul altıncı sınıf fen bilgisi dersi öğretim programındaki kazanımlara uygun 30 adet çoktan seçmeli soru hazırlanmıştır. Alan yazında test geliştirilirken madde analizi ve uzman görüşü sonrası bazı sorular çıkabileceğinden, testin taslak formunda en az üç sorunun her bir kazanımda içerilmesi gerektiği, böylece testin kapsam geçerliğinin arttırılabileceği ifade edilmektedir (Atılgan, 2013). Buna göre araştırmada her kazanımda en az üç sorunun olmasına dikkat edilmiştir. MEB (2013) FBDÖP'nda yer alan kazanımlar Tablo 1'de verilmiştir.

Tablo 1 incelendiğinde MEB (2013) FBDÖP'nda yer alan BHÜBG ünitesinin dört kazanım altında toplandığı görülmektedir. Bununla beraber daha önce de ifade edildiği gibi aynı ünite 2018 yılında güncellenen FBDÖP'nda yedinci sınıf seviyesine alt konu olarak konulmuştur. Ancak her ne kadar ders saati yeni programda azaltılmış ve sınıf seviyesi değiştirilmiş olsa da yeni

Tablo 1.
Altıncı Sınıf Fen Bilimleri Dersi Öğretim Programında BHÜBG Ünitesinin Yeri

Bölüm	F.6.5.1. Bitki ve Hayvanlarda Üreme, Büyüme ve Gelişme*
Konu/Kavramlar	Eşeysiz üreme (vejetatif üreme, bölünme, tomurcuklanma ve rejenerasyon), eşeyli üreme, büyüme ve gelişme
Kazanımlar*	F.6.5.1.1. Bitki ve hayvanlardaki üreme çeşitlerini karşılaştırır.
	F.6.5.1.2. Bitki ve hayvanlardaki büyüme ve gelişme süreçlerini örnekler vererek açıklar.
	F.6.5.1.3. Bitki ve hayvanlarda büyüme ve gelişmeye etki eden faktörleri açıklar.
	F.6.5.1.4. Bir bitki ya da hayvanın bakımını üstlenir ve gelişim sürecini rapor eder.

*Aynı kazanımlar sonradan güncellenen MEB (2018) 7. Sınıf FBDÖP'nda da yer almaktadır

Tablo 2.
Uzman Görüşleri Doğrultusunda Düzeltilen Bir Soru Örneği

Uzman görüşü öncesi	Uzman görüşü sonrası
Üreme şekli bakımından aşağıdaki hayvanlardan hangisi diğerlerinden farklıdır? A) İnek B) Koyun C) Timsah D) Maymun	Aşağıdaki hayvanlardan hangisi üreme şekli bakımından diğerlerinden farklıdır? A) İnek B) Yunus C) Timsah D) Yarasa

programın da eskisi ile aynı konu/kavramlar ve kazanımları içerdiği görülmektedir (MEB, 2018). Bu noktada bu araştırmada 2018 yılı öncesi uygulamaları yapılarak 2013 öğretim programı doğrultusunda hazırlanmış olan testin içerdiği sorular, kapsam ve kazanımlar açısından 2018 yılında güncellenen öğretim programına da uygundur.

Alan yazında yapılan araştırmalar test geliştirme süreçlerinde kapsam geçerliğine mutlaka yer verilmesi gerektiği, bunun için ise çoğunlukla uzman görüşüne başvurulduğuna vurgu yapmaktadır (Çalık & Ayas, 2003; Treagust, 1988; Tunç & Kılınç-Alpat, 2015). Uzman görüşüne başvurulması ile sadece testin kapsam geçerliği değil, aynı zamanda testin ölçülmesi istenen özelliği ölçüp ölçemeyeceği yani görünüş geçerliği de sınanabilir. Diğer taraftan uzman görüşleri testteki yazım imla hataları, eksik, hatalı veya çelişkili olan ifadelerin de düzeltilmesine katkı sağlayabilir (Çalık & Ayas, 2003). Bu duruma istinaden araştırmacılar tarafından hazırlanan 30 soruluk çoktan seçmeli taslak test, bir devlet üniversitesindeki iki biyoloji eğitimi alan uzmanı ile MEB'na bağlı bir ortaokulda görev yapan bir fen bilgisi öğretmeni tarafından incelenmiştir. Uzman görüşleri sonrası belirli kazanımda çok sayıda toplanan, benzer bilgileri ölçen, hatalı ifadeler içeren, yeni müfredata uymayan vb. nedenlerle 10 soru elenmiş ve kalan sorularda yazım, imla hataları, düşük cümleler vb. içeren az sayıda soru da düzeltilerek (Tablo 2) madde havuzu toplam 20 soruya indirilmiştir. Böylece 20 çoktan seçmeli soru içeren taslak testin madde analizleri için pilot uygulaması yapılmıştır.

Pilot uygulama sonrası toplanan veriler doğru cevaplar için "1," yanlış veya boş bırakılan cevaplar için ise "0" değeri verilerek puanlandırılmıştır. Madde analizi çalışmalarında ölçekteki maddelerin en az üç katı kadar örnekleme ulaşılması yeterli görülmektedir (Cattel, 1978). Benzer şekilde Kline (1994) ve Büyüköztürk (2005) tarafından faktör analizi, madde analizi gibi işlemlerde örneklem büyüklüğünün madde sayısının en az iki kat olması gerektiği ileri sürülmektedir. Dolayısıyla, her ne kadar bu araştırmada analiz için cevapları değerlendirmeye alınan öğrenci sayısı görünürde az olsa da alt kriteri sağlamada yeterli sayılabilir.

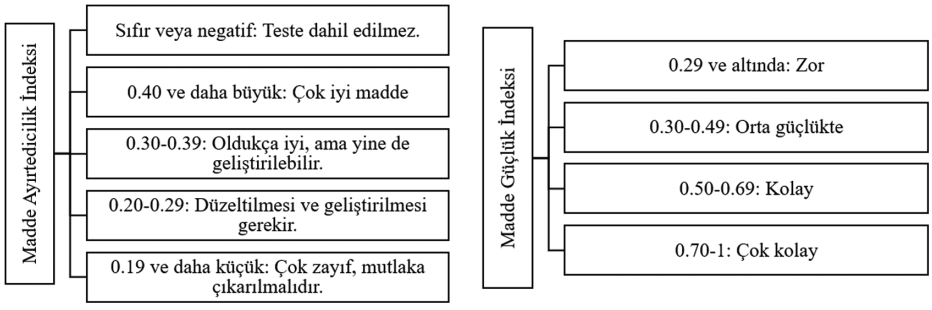
Araştırmada pilot çalışma sonrası toplanan verilerin madde analizi sürecinde madde güçlük indeksi (p) ve madde ayırt edicilik indeksi (d) değerleri hesaplanmıştır. Madde analizi sürecinde madde istatistikleri hesaplanırken alan yazında kullanılan en yaygın yöntemlerden biri Basit Yöntem'dir (Hasançebi ve ark., 2020). Basit Yöntem'de tüm yanıtlayıcıların toplam puanlarından en başarılı ve en başarısız olan %27'lik üst ve alt gruplar belirlenerek analizler yapılır. Basit Yöntem'in tercih edildiği bu araştırmada da öncelikle toplanan veriler "1" ve "0" şeklinde kodlandıktan sonra öğrencilerin testten almış oldukları puanlar en yüksek değerden başlayarak sıralanmıştır. Bu sıralama sonucunda en yüksek ve en düşük puana sahip olanlardan 15'er (55*27/100) öğrenci belirlenmiştir. Madde güçlüğü hesaplanırken $p = (Dü+Da)/2N$ formülünden, madde ayırt ediciliği hesaplanırken ise $d = (Dü-Da)/N$ formülünden yararlanılmıştır (N: Tüm grubun % 27'si, Dü: Üst grupta doğru yapanların sayısı, Da: Alt grupta doğru yapanların sayısı). Analizler sonrasında elde edilen değerler Şekil 2'de gösterilen kriterler dikkate alınarak yorumlanmıştır (Hasançebi ve ark., 2020).

Araştırmada ayrıca madde analizi sonrası son şekli verilen 20 soruluk test 60 kişilik bir örneklem grubuna uygulanarak KR-20 güvenirlik katsayısı hesaplanmıştır.

Bulgular

Araştırmada daha önce ifade edildiği gibi toplanan veriler için madde analizleri yapılmış (Tablo 3) ve elde edilen bulgular Şekil 2'de sunulan kriterler dikkate alınarak değerlendirilmiştir.

Tablo 3 incelendiğinde araştırmada kullanılan başarı testine ait 18. sorunun orta güçlükte, diğerlerinin ise kolay/çok kolay sorular olduğu anlaşılmaktadır. Bir başka ifadeyle 18. soruya ait madde güçlük indeksi 0,40 değerini aldığından orta güçlük aralığındadır. Diğer maddelere bakıldığında ise 0,50–0,69 aralığında kalan dokuz sorunun (2., 3., 4., 6., 7., 12., 15., 17. ve 20. sorular) kolay olduğu; 0,70–1,00 aralığındaki on sorunun (1., 5., 8., 9., 10., 11., 13., 14., 16. ve 19. sorular) ise çok kolay olduğu belirlenmiştir. Bununla beraber testin ortalama madde güçlüğü ,66 civarındadır. Bu ise testin kolay bir test olduğunun göstergesidir. Testin ayırt ediciliğine ait bulgular incelendiğinde ise 8, 14, 16 ve 20 numaralı soruların düzeltilmesi ve geliştirilmesi gereken sorular olduğu anlaşılmaktadır. Zira bu soruların ayırt edicilik indeks değerleri 0,20–0,29 arasında değişmektedir. Bu sorular revize edilerek testte bırakılmıştır. Analiz sonuçları incelendiğinde testte geriye kalan soruların ise 0,40 ve üzeri değer alması nedeniyle istenen düzeyde ayırt edici sorular olduğu ortaya çıkmıştır. Diğer taraftan teste ait ortalama ayırt edicilik değeri ,56 olarak belirleniştir. Buna göre testin ayırt ediciliğinin oldukça iyi olduğu ve soruların testte kullanılabileceği söylenebilir.



Şekil 2.
Madde Analizine Ait Kriterler

Tablo 3.
Testteki Sorulara Ait Madde Analizi Sonuçları

Soru no	Doğru cevap	Dü	Da	*p*	d	Soru no	Doğru cevap	Dü	Da	*P*	d
1	C	15	8	,77	,47	11	A	15	8	,77	,47
2	B	14	3	,57	,73	12	B	15	2	,57	,87
3	C	15	2	,57	,87	13	C	15	6	,70	,60
4	A	15	1	,53	,93	14	B	15	12	,90	,20
5	D	15	6	,70	,60	15	D	12	4	,53	,53
6	D	13	3	,53	,67	16	A	13	9	,73	,27
7	C	15	3	,60	,80	17	B	15	4	,63	,73
8	A	14	10	,80	,27	18	A	10	2	,40	,53
9	B	15	6	,70	,60	19	D	15	9	,80	,40
10	C	15	8	,77	,47	20	D	11	7	,60	,27

Testin geneline ait $p = ,66$
Testin geneline ait $d = ,56$

Madde analizinde ayrıca %27'lik alt ve üst dilimlerdeki puanlar arasındaki farklılığın istatistiksel olarak anlamlılığına ilişkin bağımsız örneklem t testi yapılmıştır. Yapılan t testi ile başarı testinin alt ve üst grupları yeterli düzeyde ayırt edebildiği sonucuna varılmıştır ($t = -15,558$, $p < ,05$). Madde analizlerine ek olarak güvenirlik analizi yapılmıştır. Bilindiği üzere testteki maddelere verilen cevaplar arasındaki tutarlılık, testin ölçmeyi amaçladığı özelliği ne derecede doğru ölçtüğünün göstergesi güvenirlik olarak bilinir (Büyüköztürk, 2005). 0,00 ile +1,00 arasında değerler alan güvenirlik katsayısının +1,00'e yaklaşması güvenirliğin yüksek olduğunu göstermektedir (Avcı, 2020). Çalışmada testin güvenirliğinin belirlenmesinde tek uygulamaya dayalı yöntem tercih edilmiş olup güvenirlik katsayısı (r), Kuder Richardson-20 (KR20) yöntemi ile belirlenmiştir. Buna göre SPSS istatistik programı ile yapılan analiz sonucunda testin KR-20 güvenirlik katsayısı ,93 olarak hesaplanmıştır. Bu bulgu başarı testinin oldukça güvenilir olduğunu göstermektedir.

Tablo 4.
BHÜBGBT'ne Ait Belirtke Tablosu

Kazanım	Kazanım düzeyi	Soru no	Sorunun düzeyi
6.5.1.1. Bitki ve hayvanlardaki üreme çeşitlerini karşılaştırır.	[Olgusal bilgi/ Anlama]	1	Anlama
		2	Anlama
		3	Anlama
		4	Hatırlama
		5	Anlama
		6	Anlama
		7	Hatırlama
		8	Hatırlama
		9	Anlama
6.5.1.2. Bitki ve hayvanlardaki büyüme ve gelişme süreçlerini örnekler vererek açıklar.	[Kavramsal bilgi/ Anlama]	10	Anlama
		11	Çözümleme
		12	Anlama
		13	Çözümleme
		14	Çözümleme
		15	Hatırlama
		16	Hatırlama
		17	Anlama
		18	Anlama
6.5.1.3. Bitki ve hayvanlarda büyüme ve gelişmeye etki eden faktörleri açıklar.	[Kavramsal bilgi/ Anlama]	19	Çözümleme
6.5.1.4. Bir bitki ya da hayvanın bakımını üstlenir ve gelişim sürecini rapor eder.	[İşlemsel bilgi/ uygulama]	20	Değerlendirme

*Soru numaraları testin nihai şekline göre yeniden düzenlenmiştir

Araştırmada geçerlik ve güvenirlik çalışmalarının yanı sıra testte yer alan soruların bilişsel alan düzeyindeki yeri de incelenmiştir. Bilindiği üzere 2001 yılında Anderson ve arkadaşları tarafından Bloom Taksonomisi güncellenmiştir (Anderson ve ark., 2001). Ancak alan yazında bazı araştırmalara göre yenilenmiş Bloom Taksonomisi ülkemizde sınırlı olarak kullanılmaktadır (Kaya & Gül, 2020; Tutkun, 2012). Buradan hareketle bu araştırmada BHÜBGBT'nde yer alan sorular yenilenmiş Bloom Taksonomisi açısından da incelenmiştir. Buna göre araştırmada geçerliği ve güvenirliliği sağlanan BHÜBGBT' nin yenilenmiş Bloom Taksonomisi'ne göre durumu Tablo 4'te gösterilen belirtke tablosunda gösterilmiştir.

Tablo 4 incelendiğinde, yenilenmiş Bloom Taksonomisi'nin bilişsel alan basamağına göre öğretim programında belirtilen birinci kazanıma giren dokuz sorudan üçü hatırlama, altısı anlama düzeyindedir. İkinci kazanıma giren dokuz sorudan ikisi hatırlama, dördü anlama ve üçü çözümleme basamağına girmiştir. Diğer taraftan üçüncü kazanımdaki bir soru çözümleme, dördüncü kazanımdaki bir soru ise değerlendirme basamağındadır. İlgili basamaklara giren soruların birer örneği Ek 1'de yer almaktadır.

Tartışma

Bu araştırmada altıncı sınıf öğrencilerinin fen bilimleri dersi BHÜBG konusu için öğretim programında belirtilen kazanımlar ile uyumlu, çoktan seçmeli sorulardan oluşan geçerli ve güvenilir bir başarı testi geliştirilmiştir. Bu amaçla öncelikle ilgili alan yazın taranarak öğretim programının kazanımlarına uygun 30 adet çoktan seçmeli soru hazırlanmıştır. Bununla birlikte taslak testte her kazanım için en az üç soru olmasına dikkat edilerek testin

kapsam geçerliği sağlanmaya çalışılmıştır. Öte yandan kapsam ve görünüş geçerliği çalışmalarında, 30 sorudan oluşan taslak test uzman görüşleri doğrultusunda incelenmiş, gerekli görülen durumlarda düzeltmeler yapılmış ve öğrencilerin de yaş grubu dikkate alınarak soru sayısı 20'ye indirilmiştir. Böylece 20 sorudan oluşan test, madde analizi için 55 kişilik öğrenci grubuna uygulanmıştır. Madde analizi sonucunda testin ortalama güçlüğünün 0,66, ayırt ediciliğinin ise 0,56 olduğu belirlenmiştir. Testin KR-20 güvenirlik katsayısı ise 0,93 olarak hesaplanmıştır. Alan yazında bir testin ortalama madde güçlüğünün 0,50 civarında olması gerektiği ifade edilmektedir (Bolat ve Karamustafaoğlu, 2019). Bunu destekler nitelikte birçok araştırmada geliştirilen testlerin de ortalama madde güçlük değerinin orta düzeyde olduğu ortaya konulmuştur (Aymen-Peker & Taş, 2019; Ayvacı & Durmuş, 2016; Bolat & Karamustafaoğlu, 2019; Demir ve ark., 2016). Dolayısıyla bu araştırmada geliştirilen testin madde güçlüğünün 0,66 değeri ile kolay olması arzu edilir bir durum değildir. Bununla beraber Hasançebi ve ark. (2020)'ye göre bir testte yer alan maddelerin tamamının madde güçlük indekslerinin 0,50 olacak şekilde hazırlanması tercih edilen bir durum olmayıp testte kolay, orta güçlükte veya zor olarak tanımlanabilecek soruların yer alması gerekmektedir. Buradan hareketle araştırmada çoğu kolay olan soruların bir kısmı revize edilerek testte bırakılmıştır. Diğer taraftan test geliştirme çalışmalarında madde güçlüğü açısından kolay olan testlerin de geliştirildiği görülmektedir (Açıkgöz & Karslı, 2015; Bingöl & Halisdemir, 2017; Kaya & Gül, 2020). Söz konusu araştırmalarda madde güçlük indeksinin tek başına yeterli olmadığı, madde ayırt edicilik indekslerini de incelenmesi gerektiği vurgulanmaktadır (Kaya & Gül, 2020; Kızkapan & Bektaş, 2018; Nacaroğlu ve ark., 2020). Dolayısıyla bu araştırmada madde güçlüğünün yanı sıra madde ayırt ediciliğine de bakılmıştır. Yapılan analizler ise testin ayırt ediciliğinin yüksek olduğunu göstermiştir. Bir başka ifadeyle, bu testin konuyu bilen ve bilmeyen öğrencileri ayırt etmede oldukça yeterli olduğu söylenebilir. Bu çalışmaya benzer şekilde BHÜBG konusuna yönelik Güneş ve Serdaroğlu (2020) tarafından geliştirilen testin de kolay ama yüksek düzeyde ayırtedici olduğu belirlenmiştir. Bununla beraber testler güvenirlik açısından karşılaştırıldığında bu araştırmada geliştirilen testin güvenirlik düzeyinin daha yüksek olduğu görülmektedir.

Araştırmada geliştirilen soruların kazanımlara göre dağılımına bakıldığında dokuz soru birinci kazanıma, dokuz soru ikinci kazanıma ve birer soru ise üçüncü ve dördüncü kazanıma dahil olmuştur. Buna göre soruların kazanımlara dengeli bir şekilde dağılmadığı söylenebilir. Ancak ilk iki kazanım içerik açısından daha fazla kavramı içerdiğinden çok sayıda sorunun bu kazanımlarda toplanması doğaldır. Dolayısıyla testin fen bilimleri öğretim programındaki ilgili konuya ait kazanımları ölçtüğü söylenebilir. Ayrıca sorular yenilenmiş Bloom Taksonomisi açısından incelendiğinde, beş soru hatırlama, on soru anlama, dört soru çözümleme ve bir soru değerlendirme basamağına girmiştir. Buna göre testin bilişsel alanın üst basamaklarına uygun sorular içerdiği söylenebilir. Daha önce de ifade edildiği gibi aynı konuya yönelik bir başarı testi Güneş ve Serdaroğlu (2020) tarafından da geliştirilmiştir. Ancak ilgili çalışmaya bakıldığında geliştirilen 40 soruluk testin kazanımlara dağılımı ve Bloom Taksonomisi açısından durumu ile ilgili bir bilgilendirme bulunmadığından diğer çalışmaya benzer veya farklı yönleri bu açıdan karşılaştırılamamıştır. Bununla beraber bu araştırmada geliştirilen test 20 soru içerdiğinden öğretmenler tarafından konu sonu kısa süreli sınav olarak kullanılabileceği gibi akademisyenler tarafından eğitim araştırmalarında, özellikle deneysel çalışmalarda kullanılabilir. Diğer taraftan bilindiği üzere gerek bu araştırmada gerekse Güneş ve Serdaroğlu (2020) tarafından yapılan araştırmada BHÜBG konusuna yönelik geliştirilen başarı testleri 2013 yılı öğretim programına göre yapılmış olmakla birlikte, 2018 yılında güncellenen yeni FBDÖP'nda her ne kadar ders süresi 12 saate indirilmiş olsa da aynı kazanımların yer alması nedeniyle bu araştırmada geliştirilen başarı testinin güncel programa da uygun olduğu düşünülmektedir.

Sonuç ve Öneriler

Sonuç olarak yukarıda geliştirilme aşamaları verilen BHÜBGBT'nin gerek 2013 gerekse 2018 yılında güncellenen yeni fen bilimleri öğretim programı kazanımları ile uyumlu olduğu ve programda yer alan içeriği tamamen kapsadığı görülmektedir. Bununla birlikte araştırma sürecinde karşılaşılan birtakım durumlar dikkate alındığında aşağıdaki önerilerin gelecek araştırmalar için yol gösterici olacağına inanılmaktadır. Buna göre:

1. Araştırmada başarı testi, madde analizi ve güvenirlik analizleri gibi istatistiksel yöntemler kullanılarak son şeklini almıştır. Bununla beraber, son yıllarda yapılan bazı araştırmalarda başarı testlerinin geliştirilmesi sürecinde faktör analizi yöntemi ile yapı geçerliğine de bakılması gerektiği ileri sürülmektedir (Demir ve ark., 2016; Keçeci ve ark., 2019; Nacaroğlu ve ark., 2020). Bilindiği üzere başarı testilerinin puanlanması çoğunlukla doğru cevaplar için 1, yanlış/boş bırakılan cevaplar için 0 şeklinde kodlanarak yapılmaktadır. Bu nedenle özellikle anketlerde sıklıkla tercih edilen ve çoğunlukla 1–5 arası kodlamaların yapıldığı puanlamalarda çoğunlukla SPSS programı yardımıyla tipik faktör analizi yapılabilmektedir. Halbuki başarı testlerinde 1-0 şeklinde kodlanan puanlar için bilinen istatistik programlarından (SPSS gibi) farklı olarak Mplus, FACTOR gibi programlar kullanılmak olup bu programlarda tetrakorik faktör analizi yapılarak testin yapı geçerliği incelenmektedir. Tetrakorik korelasyon katsayısı, 1-0 gibi iki kategorili yapay süreksiz iki değişken arasındaki ilişkinin derecesini belirlemek için kullanılmaktadır (Baykul & Güzeller, 2014; Dokumacı-Sütçü & Oral, 2019). Bu araştırmada ise ilgili programların kullanımı ve tetrakorik faktör analizi süreci araştırmacıların uzmanlık alanı dışında olduğundan yapılamamıştır. Bu nedenle yapı geçerliğinin olmaması bu araştırmanın bir eksikliği olarak nitelendirilebilir. Buradan hareketle ileride bu konuya yönelik yapılacak başarı testi geliştirme çalışmalarında tetrakorik faktör analizi yöntemi ile testin yapı geçerliğinin incelenmesi önerilmektedir.
2. Araştırmada her ne kadar alan yazında test geliştirme sürecinde olması gereken örneklem sayısına ulaşılsa da daha da güvenilir sonuçlar elde etmek adına daha yüksek sayıda örneklemlerle çalışılarak yeni başarı testleri geliştirilebilir.
3. Farklı okullarda başarı düzeyi açısından farklı örneklemler seçilerek benzer bir test geliştirilip, testin daha geniş kitlelere hitap etmesi ve sonuçlarının genellenebilirliği sağlanabilir.
4. Gelecek araştırmalarda aynı konuda başarı testleri geliştirilirken açık uçlu sorular da eklenerek testin üst düzey becerileri ölçmedeki yeterliliği daha da arttırılabilir. Özellikle çalışmada geliştirilen test çoktan seçmeli sorulardan oluştuğundan Bloom Taksonomisi'nin üst düzey basamaklarına hitap eden sorular içerme noktasında yetersiz kalmıştır. Taksonominin hem daha üst düzey basamakların hitap eden hem de kazanımlara da dengeli dağılım gösteren sorular içeren yeni başarı testleri geliştirilebilir.

5. Pilot uygulama yapılarak öğrencilerin bu konudaki başarı düzeylerini ölçmede testin gücü ölçülebilir.
6. Araştırmada öğrencilerin BHÜBG konusuna yönelik bilgi düzeylerini ölçemeye yönelik test geliştirilmiştir. Bununla beraber ileride yapılacak araştırmalarda aynı konuya ilişkin iki veya üç aşamalı bir kavram yanılgısı testi geliştirilebilir.

Etik Komite Onayı: Bu çalışma için etik komitesi onayı Atatürk Üniversitesi Sosyal ve Beşeri Bilimler Etik Kurulundan (Tarih: 04.09.2014, No: 48553601-5616) alınmıştır.

Hakem Değerlendirmesi: Dış başımsız.

Yazar Katkıları: Fikir - M.S.; Tasarım - F.Y., Ş.G., M.S.; Denetleme - F.Y., Ş.G., M.S.; Kaynaklar - F.Y, Ş.G.; Veri Toplanması ve İşlemesi - F.Y., Ş.G.; Analiz ve Yorum - Ş.G.; Literatür Taraması - F.Y., Ş.G.; Eleştirel İnceleme - F.Y., Ş.G., M.S.

Teşekkür: Bu çalışma TÜBİTAK tarafından 114K725 nolu proje kapsamında desteklenmiştir. Araştırmaya gönüllü olarak katılan öğretmen ve öğrencilere teşekkür ederiz.

Çıkar Çatışması: Yazarlar, çıkar çatışması olmadığını beyan etmiştir.

Finansal Destek: Bu çalışma TÜBİTAK tarafından 114K725 no ve 362.168 TL ile desteklenen bir projeden üretilmiştir.

Ethics Committee Approval: Ethics committee approval was received for this study from Atatürk University Social and Human Sciences Ethics Committee (Date: 04.09.2014, No: 48553601-5616).

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – M.S.; Design – F.Y., Ş.G., M.S.; Supervision – F.Y., Ş.G., M.S.; Materials – F.Y, Ş.G.; Data Collection and/or Processing – F.Y, Ş.G.; Analysis and/or Interpretation – Ş.G.; Literature Review – F.Y, Ş.G.; Critical Review – F.Y., Ş.G., M.S.

Acknowledgments: This work was supported by TUBITAK within the scope of project 114K725. We thank the teachers and students who voluntarily participated in the study.

Declaration of Interests: The authors declare that they have no competing interest.

Funding: This study was produced from a project supported by TUBITAK with the number 114K725 and 362,168 TL.

Kaynaklar

Açıkgöz, M., & Karslı, F. (2015). Alternatif ölçme-değerlendirme yaklaşımları kullanılarak iş ve enerji konusunda geliştirilen başarı testinin geçerlilik ve güvenirlik analizi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 4(1), 1–25.

Akandere, M., Özyalvaç, N., & Duman, S. (2010). Ortaöğretim öğrencilerinin beden eğitimi dersine yönelik tutumları ile akademik başarı motivasyonlarının incelenmesi (Konya Anadolu Lisesi örneği). *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 24, 1–10.

Akbulut, H. İ., & Çepni, S. (2013). Bir üniteye yönelik başarı testi nasıl geliştirilir?: İlköğretim 7. sınıf kuvvet ve hareket ünitesi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 2(1), 18–44.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., & Wittrock, M. C. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives (Complete Edition)*. Longman.

Arslan, A. (2007). *Fen Eğitiminde Araştırmaya Dayalı Öğretim Yönteminin Kavramsal Öğrenmeye Etkisi* (Yüksek Lisans Tezi), Marmara Üniversitesi, Yükseköğretim Kurulu Ulusal Tez Merkezi (Tez No. 221514).

Atılgan, H. (2013). *Eğitimde Ölçme ve Değerlendirme*. (6. Baskı), Anı Yayıncılık.

Atılgan, H., Kan, A., & Doğan, N. (2011). *Eğitimde Ölçme ve Değerlendirme*. (5. Baskı), Anı Yayıncılık.

Avcı, F. (2020). Yenilenmiş Bloom taksonomisine göre madde ve ısı başarı testi: geçerlik güvenirlik çalışması. *Iğdır Üniversitesi Sosyal Bilimler Dergisi*, 21, 263–292.

Aymen Peker, E., & Taş, E. (2019). Üç aşamalı kavram tanı testi geliştirme ve geçerlik çalışması: 5. sınıf canlılar dünyasını gezelim tanıyalım ünitesi. *Journal of Computer and Education Research*, 7(14), 515–539. [CrossRef]

Ayvacı, H. Ş., & Durmuş, A. (2016). Bir başarı testi geliştirme çalışması: ısı ve sıcaklık başarı testi geçerlik ve güvenirlik araştırması. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 35(1), 87–103.

Baykul, Y., & Güzeller, C. O. (2014). Sosyal Bilimler İçin İstatistik: SPSS Uygulamalı. Pegem Akademi Yayıncılık.

Bingöl, A., & Halisdemir, N. (2017). Üniversite öğrencilerinin temel bilgi teknolojileri dersine yönelik akademik başarı testi geliştirme çalışması. *Journal of Academic Social Science Studies*, 54, 541–554. [CrossRef]

Bolat, A., & Karamustafaoğlu, S. (2019). Vücudumuzdaki sistemler ünitesi başarı testi geliştirme: geçerlik ve güvenirlik. *Gazi Eğitim Bilimleri Dergisi*, 5(2), 131–159. [CrossRef]

Büyüköztürk, Ş. (2005). Anket geliştirme. *Türk Eğitim Bilimleri Dergisi*, 3(2), 133–151.

Çalık, M., & Ayas, A. (2003). Çözeltilerde kavram başarı testi hazırlama ve uygulama. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 2(14), 1–17.

Çardak, Ç. S., & Selvi, K. (2018). Öğretim ilke ve yöntemleri dersi için bir başarı testi geliştirme süreci. *Akdeniz Eğitim Araştırmaları Dergisi*, 12(26), 379–406. [CrossRef]

Cattell, R. B. (1978). *The Scientific Use of Factor Analysis in Behavioral and Life Sciences*. Plenum Press.

Çepni, S., Bayrakçeken, S., Yılmaz, A., Yücel, C., Semerci, Ç., Köse, E., Sezgin, F., Demircioğlu, G., & Gündoğdu, K. (2009). *Ölçme ve Değerlendirme*. (3. Baskı), Pegem Yayıncılık.

Demir, N., Kızılay, E., & Bektaş, O. (2016). 7. Sınıf çözeltiler konusunda başarı testi geliştirme: geçerlik ve güvenirlik çalışması. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 10(1), 209–237. [CrossRef]

Demirci, N., & Efe, S. (2007). İlköğretim öğrencilerinin ses konusundaki kavram yanılgılarının belirlenmesi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 1(1), 23–56.

Dokumacı Sütçü, N., & Oral, B. (2019). Uzamsal görselleştirme testinin geliştirilmesi: Geçerlik ve güvenirlik çalışmaları. *Kastamonu Eğitim Dergisi*, 27(3), 1179–1195.

Duruk, U., Akgün, A., Doğan, C., & Gülsuyu, F. (2017). Examining the learning outcomes included in the Turkish science curriculum in terms of science process skills: A document analysis with standards-based assessment. *International Journal of Environmental and Science Education*, 1(2), 117–142.

Gül, Ş. (2019). Yaşam Temelli Biyoloji Motivasyon Ölçeği (YTBMÖ) geliştirilmesi: Geçerlik ve güvenirlik çalışması. *İnönü Üniversitesi Eğitim Bilimleri Enstitüsü Dergisi*, 6(11), 63–77. [CrossRef]

Güneş, M. H., & Serdaroğlu, C. (2020). Validity and reliability of achievement test developed in reproduction, growth and development in plants and animals. *Science, Education, Art and Technology Journal (SEAT Journal)*, 2(1), 35–40.

Güngörmez, H., & Akgün, A. (2018). Ortaokul öğrencilerinin fen bilimleri dersindeki kuvvet ve enerji ünitesine yönelik akademik başarı testi geliştirme çalışması. *Diyalektolog Ulusal Sosyal Bilimler Dergisi*, 18, 85–99. [CrossRef]

Haladyna, T. M. (1997). *Writing Test Items to Evaluate Higher Order Thinking*. Allyn & Bacon.

Haras, Ö. (2009). *"Üreme" Ünitesinin 5E Modeline Göre Öğretiminin Öğrencilerin Kavramsal Anlama ve Tutumları Üzerine Etkisi.* (Yüksek Lisans Tezi), Dokuz Eylül Üniversitesi (Tez No. 239330).

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

İlhan, N., & Hoşgören, G. (2015). Fen bilimleri dersine yönelik yaşam temelli başarı testi geliştirilmesi: asit baz konusu. *Fen Bilimleri Öğretimi Dergisi*, 5(2), 87–110.

Kaya, S., & Gül, Ş. (2020). 11. sınıflar için 'Sindirim Sistemi' konusuna yönelik başarı testi geliştirme çalışması. *Uluslararası Sosyal ve Eğitim Bilimleri Dergisi*, 7(13), 72–97. [CrossRef]

Keçeci, G., Yıldırım, P., & Kırbağ Zengin, F. (2019). Sistemler akademik başarı testi: geçerlik ve güvenirlik çalışması. *Ulusal Eğitim Akademisi Dergisi*, 3(1), 96–114. [CrossRef]

Kızkapan, O., & Bektaş, O. (2018). Fen eğitiminde başarı testi geliştirilmesi: hücre bölünmesi ve kalıtım örneği. *Uluslararası Eğitim Bilimleri Dergisi*, 2(1), 1–18.

Kline, P. (1994). *An Easy Guide to Factor Analysis*. Routledge.

Küçükahmet, L. (2004). *Öğretimde Planlama Değerlendirme*. Nobel Yayın Dağıtım.

McMillan, J. H., & Schumacher, S. (2010). *Research in Education: Evidence-Based Inquiry* (7th ed). Pearson.

MEB (Millî Eğitim Bakanlığı). (2013). *İlköğretim Kurumları (Ilkokullar ve Ortaokullar) Fen Bilimleri Dersi (3, 4, 5, 6, 7 ve 8. Sınıflar) Öğretim Programı*. MEB.

MEB (Millî Eğitim Bakanlığı). (2018). *İlköğretim Kurumları (Ilkokullar ve Ortaokullar) Fen Bilimleri Dersi (3, 4, 5, 6, 7 ve 8. Sınıflar) Öğretim Programı*. MEB.

Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2001). Assessing understanding in biology. *Journal of Biological Education*, 35(3), 118–124. [CrossRef]

Nacaroğlu, O., Bektaş, O., & Kızkapan, O. (2020). Madde döngüleri ve çevre sorunları konusunda başarı testi geliştirme: geçerlik ve güvenirlik çalışması. *Kastamonu Eğitim Dergisi*, 28(1), 36–51. [CrossRef]

Önder, K. (2011). *İlköğretim 6. Sınıf Fen ve Teknoloji Dersi "Canlılarda Üreme, Büyüme ve Gelişme" Ünitesinin Öğretiminde Laboratuvar Yönteminin Öğrenci Başarısına Etkisi* (Yüksek Lisans Tezi), Selçuk Üniversitesi (Tez No. 199578).

Özcan, H., Koca, E., & Söğüt, M. (2019). Ortaokul öğrencilerinin basınç kavramıyla ilgili anlayışlarını ölçmeye yönelik bir test geliştirme çalışması. *Eğitim ve Toplum Araştırmaları Dergisi*, 6(1), 130–144.

Şahin, Ç., Bülbül, E., & Durukan, Ü. G. (2013). Öğrencilerin gök cisimleri konusundaki alternatif kavramlarının giderilmesinde kavramsal değişim metinlerinin etkisi. *Bilgisayar ve Eğitim Araştırmaları Dergisi*, 1(2), 38–64.

Saraç, H. (2018). Fen bilimleri dersi 'maddenin değişimi' ünitesi ile ilgili başarı testi geliştirme: geçerlik ve güvenirlik çalışması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 18(1), 416–445. [CrossRef]

Tekbıyık, A., & Akdeniz, A. R. (2010). Bağlam temelli ve geleneksel fizik problemlerinin karşılaştırılması üzerine bir inceleme. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 4(1), 123–140.

Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, 10(2), 159–169. [CrossRef]

Tunç, T., & Kılınç-Alpat, S. (2015). Elektrokimyada geçerlik ve güvenirliği sağlanmış kimya başarı testinin geliştirilmesi. *Yükseköğretim ve Bilim Dergisi*, 5(3), 347–360.

Tutkun, Ö. F. (2012). Bloom'un yenilenmiş taksonomisi üzerine genel bir bakış. *Sakarya University Journal of Education*, 1(3), 14–22.

Ürey, M., & Cerrah-Özsevgeç, L. (2015). Sınıf öğretmen adaylarının fen bilgilerini günlük yaşamla ilişkilendirebilme düzeyleri ile fen tutum ve okuryazarlıkları arasındaki ilişki. *Kuramsal Eğitimbilim Dergisi*, 8(3), 397–420.

Yip, D. Y. (1998). Children's misconceptions on reproduction and implications for teaching. *Journal of Biological Education*, 33(1), 21–26. [CrossRef]

Extended Abstract

Purpose: Today, the main purpose of science education given to students is not to provide them with the current knowledge but to enable them to learn the ways of accessing it. In this way, an individual who learns by doing and experiencing can both develop scientific process skills and solve problems related to new situations in daily life. It is possible to determine how much the students understand the information transferred to them by measuring only their achievement. Academic achievement and failure levels of the students in the learning environment can be determined by measurement and assesment. In the assessment process, it is tried to determine whether each student has a certain feature or qualification or not. In this process, teachers firstly set certain goals in their courses and create appropriate teaching situations to achieve these goals. They determine the realization level of the goals by using the measurement results. Qualified measurement tools are needed for this determination to be healthy. The different types of tests are generally used as measuring instruments. For example, when the literature is examined, it is seen that success tests are mostly used in studies where students' success is examined as dependent variable.

Multiple choice tests, which are used to reveal the missing and incorrect learning of the students, are more useful than other measurement tools in terms of time and preparation and can be easily applied in the classroom environment. However, it is extremely important that the success tests, which are prepared as multiple-choice, must be prepared with the appropriate analysis and methods by ensuring their validity and reliability.

There are many studies to develop success test in the field of science education. When we look at the success test development studies on biology which is an important field of science, it is remarkable that it is quite limited. In this study, a success test related to "Reproduction, Growth and Development in Plants and Animals" subject was prepared. The purpose of selecting the "Reproduction, Growth and Development in Plants and Animals" subject is to make it the firstly place in a research on the difficulty of understanding secondary school students' biology subjects in science course. As result, the main purpose of the study is to develop a achievement test to measure the success of sixth grade students in the "Reproduction, Growth and Development in Plants and Animals" subject.

Method: In this research, in which the quantitative research paradigm was adopted, the survey method was used. Survey method is used to learn about people's attitudes, beliefs, values, demographics, behavior, opinions, habits, desires and other types of information.

Result: This survey research was carried out in several steps. Firstly, the researchers examined the literature related to the purpose and then they identified a total of 30 multiple choice questions. In the determination of the questions, the learning outcomes in the science curriculum were taken into consideration. For the validity of the test, an expert opinion was applied, so that language, scope, comprehensibility, etc., were presented to the opinion of academicians and teachers who are experts in their fields. After the expert opinions, the questions were reduced to 20 items. The 20-question test was applied to a total of 55 students who had previously studied the subject for pilot applications. The data obtained were subjected to item analysis and editing was performed. In the item analysis, the average difficulty of the test consisting of 20 multiple choice questions was calculated as 0.66 and the average discriminant was 0.56. After the item analysis, some of the remaining questions were corrected. Then, the reliability test was applied to a different group of 60 students for reliability analysis and the KR-20 reliability coefficient was calculated as 0.93.

Discussion and Conclusion: When the findings are evaluated in general, it can be said that the test is a valid and reliable test with questions of different difficulty levels. However, this research has some limitations. For example, students' high level skills can be measured by adding different types of questions (eg. open ended) in achievement tests developed in future studies. On the other hand, in the research, a test was developed to measure the knowledge level of the students about the subject. A misconception test on the same subject can be developed in future research. Additionally, two or three-stage multiple-choice tests can be developed for this purpose. Finally, it may be suggested that more than one application of the draft form of the test to be developed by selecting schools with different success levels. Structural validity of the test can be examined by performing tetrachoric factor analysis.