

ÖZEL YETENEKLİ ÖĞRENCİLER İÇİN SÜRDÜRÜLEBİLİRLİK ÜNİTESİNE YÖNELİK AKADEMİK BAŞARI TESTİ: GEÇERLİK VE GÜVENİRLİK ÇALIŞMASI*

Academic Achievement Test for The Sustainability Unit for Gifted Students: Validity And Reliability Study*

Sait Burak YILMAZ¹; Yavuz YAMAN²

ÖZ

Geliş Tarihi

10.12.2025

Kabul Tarihi

07.04.2026

Yayın Tarihi

30.06.2026

Anahtar Kelimeler

Başarı Testi,
Özel Yetenekliler,
Sürdürülebilirlik,
Ölçme ve
Değerlendirme.

Çalışmanın amacı, dördüncü ve beşinci sınıfta öğrenim gören özel yetenekli öğrenciler için “Sürdürülebilirlik” ünitesine yönelik geçerli ve güvenilir bir akademik başarı testi geliştirmektir. Bu çalışma, bir test geliştirme ve psikometrik değerlendirme çalışması olarak yürütülmüştür. Çalışmanın örneklemini, 2024–2025 eğitim öğretim yılında Türkiye’nin farklı bölgelerindeki Bilim ve Sanat Merkezlerinde öğrenim gören ve ölçüt örnekleme yöntemiyle seçilen toplam 238 öğrenci oluşturmaktadır. Araştırmada veri toplama aracı olarak “Sürdürülebilirlik” ünitesi akademik başarı testi taslak formu kullanılmıştır. Kapsam ve görünüş geçerliği için belirtke tablosu hazırlanarak uzman görüşü alınmıştır. Çoktan seçmeli maddelerin analizinde Test Analysis Program (TAP), açık uçlu maddelerin analizinde ise puanlayıcılar arası tutarlılığı belirlemek amacıyla Intraclass Correlation Coefficient (ICC) kullanılmıştır. Geliştirilen başarı testi; dört seçenekli 29 çoktan seçmeli madde ile 9 açık uçlu maddeden oluşmaktadır. Çoktan seçmeli maddeler için ortalama madde güçlük indeksi 0.55, madde ayırt edicilik indeksi 0.47 olarak hesaplanmıştır. Testin güvenirlik katsayıları KR-20 için 0.81, KR-21 için 0.80’dir. Açık uçlu maddelerde madde güçlük indekslerinin 0.22–0.62, ayırt edicilik indekslerinin 0.35–0.72 aralığında değiştiği; puanlayıcılar arası tutarlılığın ise maddeler bazında ICC = 0.98, toplam puan bazında ICC = 0.99 olduğu belirlenmiştir.

ABSTRACT

Submit Date

10.12.2025

Accept Date

07.04.2026

Publish Date

30.06.2026

Keywords

Achievement Test,
Gifted Students,
Sustainability,
Assessment and
Evaluation

The purpose of this study is to develop a valid and reliable academic achievement test for the sustainability unit for gifted students in the fourth and fifth grades. This study was conducted as a test development and psychometric evaluation study. The sample of the study consisted of a total of 238 students selected by criterion sampling who were studying at Science and Art Centers in different regions of Turkey in the 2024–2025 academic year. The draft form of the Sustainability Unit Academic Achievement Test was used as the data collection tool in the research. A table of specifications was prepared and expert opinion was obtained for content and face validity. The Test Analysis Program (TAP) was used to analyze the multiple-choice items, while the Intraclass Correlation Coefficient (ICC) was used to determine inter-rater reliability for the open-ended items. The developed achievement test consists of 29 multiple-choice items with four options and 9 open-ended items. The average item difficulty index for multiple-choice items was calculated as 0.55, and the item discrimination index as 0.47. The reliability coefficients of the test are 0.81 for KR-20 and 0.80 for KR-21. For the open-ended items, item difficulty indices ranged from 0.22 to 0.62, discrimination indices ranged from 0.35 to 0.72, and inter-rater reliability was ICC = 0.98 at the item level and ICC = 0.99 for total scores.

* Bu araştırma birinci yazarın, ikinci yazar danışmanlığında hazırladığı doktora tezinden üretilmiştir.

¹  Arş. Gör., İstanbul Üniversitesi-Cerrahpaşa, Lisansüstü Eğitim Enstitüsü, Türkiye, saitburak.yilmaz@ogr.iuc.edu.tr

²  Dr. Öğr. Üyesi, İstanbul Üniversitesi-Cerrahpaşa, Hasan Ali Yücel Eğitim Fakültesi, Türkiye, yyaman@iuc.edu.tr

© Yılmaz, S.B.; Yaman, Y. Bu çalışmanın telif hakları yazarlara aittir.



Bu çalışma [Creative Commons Attribution \(CC-BY\) 4.0 Lisansı](https://creativecommons.org/licenses/by/4.0/) kapsamında lisanslanmıştır.
This study is licensed under a [Creative Commons Attribution \(CC-BY\) 4.0 License](https://creativecommons.org/licenses/by/4.0/)

1. Giriş

21. yüzyıl; dijital dönüşümün ve küresel etkileşimin çok hızlı bir şekilde gerçekleştiği bireylerin bilgi edinme, edindiği bilgiyi kullanma ve öğrenme şekillerini bu yönde dönüştürdüğü bir dönemdir. Bilgiye erişimin sınırlarının büyük ölçüde ortadan kalkması ve yaşam boyu öğrenme sorumluluğunun artması, eğitimin geleneksel bakış açısı ve felsefesiyle sürdürülemediğini ortaya çıkarmaktadır (Leu vd., 2013; Organisation for Economic Cooperation and Development [OECD], 2018). Bu durum öğretim programlarının temel bileşenlerinden hedef, içerik, öğrenme-öğretme süreci ile ölçme ve değerlendirme boyutlarının yeniden yapılandırılmasını sağlarken aynı zamanda; öğrenmenin niteliğinin, öğrenenlerin sürece katılım şekillerinin ve öğrenme çıktılarının nasıl değerlendirileceğinin yeniden gözden geçirilmesine neden olmuştur. Süreçte ortaya çıkan değerlendirme ihtiyacı, öğrencilerin gerçek yaşamla ilişkili, çok boyutlu ve disiplinler arası problem alanlarıyla etkileşimini arttıran tema ve içeriklerin öğretim programlarında daha belirgin şekilde yer almasını sağlamıştır. Sürdürülebilirlik konusu da öne çıkan bu alanların başında gelmektedir.

Sürdürülebilirlik ve çevre ile ilgili konular söz konusu dönüşüm sürecinde öğrencilerin gerçek yaşam problemlerini disiplinler arası bir bakış açısı ile ele almasını destekleyen önemli öğrenme alanlarından biridir. Özellikle özel yetenekli öğrenciler açısından bu alan hem kendi yaşamları hem de içinde bulundukları toplum ile öğretim programı arasındaki bağlantıları daha kolay şekilde fark etmelerine olanak sağlayabilir (Corash ve Jones, 2012). Söz konusu konuların doğası gereği farklı disiplinleri bir arada barındırması ve açık uçlu bir yapıya sahip olması özel yetenekli öğrenciler için oldukça uygun bir öğrenme alanı oluşturabilir (Cross, 2011). Nitekim bu öğrenciler, karşılaştıkları yerel ya da küresel problem durumlarına akranlarına göre daha özgün çözüm önerileri geliştirme eğilimindedir (Silverman, 2003). Bu durum özel yetenekli öğrencilerin sahip oldukları hazırbulunuşluk düzeyinin çoğu zaman öğretmenlerinin öngördüğünden daha yüksek olabildiğini göstermektedir. Özel yetenekli öğrencilerin sahip oldukları bu yüksek bilişsel ve duyuşsal kapasite çoğu zaman aynı yaş grubundaki akranlarıyla benzer gelişimsel dönemde bulunmalarına karşın standart sınıf uygulamalarının ihtiyaçlarını karşılamada yetersiz kalmasına neden olabilmektedir. Sonuç olarak, eğitimsel standartların ve normların üzerinde farklı öğrenme gereksinimleri bulunan bu öğrenciler için öğretim programlarının yeniden yapılandırılması önem kazanmaktadır (McCollister ve Sayler, 2010). Yeniden yapılandırılan öğretim programlarının etkililiği, programın hedeflediği öğrenme çıktılarıyla uyumlu ölçme araçları kullanılarak değerlendirildiğinde daha açık şekilde ortaya konulabilir. Özellikle akıl yürütme, analiz ve değerlendirme gibi üst düzey bilişsel süreçleri geliştirmeyi amaçlayan farklılaştırılmış uygulamalarda, bu süreçleri yansıtabilen ölçme araçlarının geliştirilmesi önemlidir. Türkiye bağlamında da özel yetenekli öğrencilerle yürütülen farklılaştırılmış öğretim uygulamalarının değerlendirilmesinde benzer bir ölçme aracının geliştirilmesi ihtiyacı ortaya çıkmaktadır.

Özel yetenekli öğrencilerin akademik başarılarının değerlendirilmesinde, kendi sınıf düzeylerine uygun standart başarı testleri ve yetenek testleri her zaman yeterli ölçme duyarlılığı sunmayabilir. Sınıf düzeyi testler çoğunlukla ortalama öğrenci performansını ölçmek amacıyla geliştirildiği için özel yetenekli öğrencilerin sahip oldukları gerçek yeterlikleri ya da potansiyeli ayırt etmede sınırlı kalabilmektedir. Nitekim testin en yüksek puan sınırının (tavanının) düşük olması, özel yetenekli öğrencilerin puanlarının birbirine çok yakın değerlerde toplanmasına neden olabilmekte; bu durum öğrenciler arasındaki gerçek farklılıkların görünürlüğünü azaltarak yapılan istatistiksel yorumları da zayıflatılabilmektedir (Warne, 2012). Alan yazında yaygın kabul gören yaklaşımlardan biri, daha üst yaş ya da sınıf düzeyi için hazırlanan testlerin özel yetenekli öğrencilere uygulanmasıdır. Bu yaklaşımın temel gerekçeleri arasında test tavanını yükseltme, ayırt ediciliği artırma ve güvenilirlik gibi psikometrik özellikleri güçlendirme durumları yer almaktadır. Geliştirilen birçok başarı testinde ortaya çıkan tavan etkisi durumunun özel yetenekli öğrencilerin öğrenme kazanımlarını izlemeyi güçleştirdiği ve program değerlendirmelerinde yaşanan bazı sorunların ölçme güçlükleriyle ilişkili olabildiği de vurgulanmaktadır (Warne, 2014). Bu çerçevede, geliştirilen ölçme aracı ile “Sürdürülebilirlik” ünitesi kapsamında özel yetenekli öğrencilerin başarı düzeyini daha duyarlı şekilde belirlemeye yönelik geçerlik ve güvenilirlik kanıtlarının üretilmesi amaçlanmaktadır.

2024 yılında Millî Eğitim Bakanlığı [MEB] tarafından geliştirilen Türkiye Yüzyılı Maarif Modeli kapsamında Bilim ve Sanat Merkezleri’ne yönelik hazırlanan fen bilgisi dersi çerçeve öğretim programına “Sürdürülebilirlik” ünitesi ve alt temaları eklenmiştir (MEB, 2024). Programdaki bu değişiklik dördüncü ve beşinci sınıf düzeyindeki özel yetenekli öğrencilerin sürdürülebilirlik konusundaki akademik başarılarının ve farklı bilişsel süreç becerilerinin uygun ölçme araçlarıyla değerlendirilmesine yönelik gereksinimi artırmıştır. Bu doğrultuda geliştirilecek başarı testi öğrencilerin

sürdürülebilirlik konusuna ilişkin bilgi düzeylerinin ortaya konulmasının yanı sıra süreçte gerçekleştirilen öğretimsel uygulamaların etkililiğinin belirlenmesine de katkı sağlayabilir.

Günümüzde ölçme ve değerlendirme yaklaşımları öğrenme sürecini yalnızca öğrencilerin bilgi düzeylerini belirleyen bir uygulama olmaktan çıkararak; öğrencilerin bilgiyi nasıl yapılandırdıklarını, hangi düşünme süreçlerini kullandıklarını ve bilgiyi farklı bağlamlara transfer edebilme becerilerini ortaya koyan çok boyutlu bir değerlendirme sistemine dönüştürmüştür (Brookhart, 2010; Laverty vd., 2016). Bu dönüşüm eğitim çıktılarının daha bütüncül şekilde izlenmesine olanak sağlayarak öğrenci performansının geçerli ve güvenilir yöntemlerle değerlendirilebilmesine yönelik yeni ölçme araçlarının geliştirilmesine de olanak sağlamıştır. Eğitimde kullanılan ölçme araçları; bilişsel, duyuşsal ve devinışsel alanlara yönelik farklılık gösterse de özellikle bilişsel alan davranışlarının belirlenmesinde çoktan seçmeli testler hem ulusal sınavlarda hem de sınıf içi değerlendirme uygulamalarında en yaygın kullanılan yöntemlerden biri olmuştur (Karamustafaoğlu ve Tutar, 2016).

Başarı testlerinde kullanılan madde türleri; ölçülmek istenen öğrenme çıktılarının niteliğine göre avantaj ve sınırlılıklar taşımaktadır (Birgin ve Özcan, 2022). Çoktan seçmeli testler geçerlik, güvenilirlik ve kullanılabilirlik açısından eğitimciler için önemli avantajlar sunmaktadır; ancak şans başarısının ölçme sonuçlarını etkileyebilmesi, etkili çeldiriciler yazmanın güçlüğü ve analiz ile değerlendirme gibi üst düzey bilişsel hedeflere yönelik madde yazımındaki zorluklar bu test türünün başlıca sınırlılıkları arasında yer almaktadır (Atalmış, 2022). Çoktan seçmeli testler, öğrencilerin kendilerini ifade etme ve yaratıcı düşüncelerini ortaya koyma fırsatlarını sınırlandırabilir (Çetin, 2010). Bu duruma karşılık açık uçlu maddeler öğrencilerin akıl yürütme ve yorumlama becerilerini ortaya koymaları için fırsat sunduğundan üst düzey bilişsel becerilerin ölçülmesinde daha işlevsel olabilir (Popham, 2003). Bu tür maddeler öğrencilerin tek bir doğru cevabı olmayan ve bazı bilgilerin eksik bırakıldığı gerçek yaşam problemlerini çözmelerini gerektirebilir. Bu süreçte öğrencilerden yalnızca sorunu çözmek amacıyla bilgiyi kullanmaları değil, aynı zamanda problemi ele alma ve çözüm sürecine ilişkin düşünceleri beklenmektedir. Eksik verilen bilgilerle ilgili kendi çıkarımlarını yapmaları ve durumu yorumlayarak yaratıcı çözümler geliştirmeleri de sağlanabilir (Akay vd., 2006). Öğrencilerin üst düzey düşünme becerilerinin ölçülmesinde açık uçlu maddelerin kullanımı yaygınlaşmakla birlikte açık uçlu testlerde puanlama süreci, değerlendiricinin yorumuna bağlı olduğundan aynı yanıt farklı puanlayıcılar tarafından farklı şekillerde değerlendirilebilir (Karakaya ve Şata, 2022).

Açık uçlu testlerde puanlama sürecinin değerlendirici yorumuna bağlı olması, ölçme sonuçlarının tutarlılığını olumsuz etkileyebilmekte ve uygulamada daha nesnel bir değerlendirme biçimi sunan çoktan seçmeli testlere yönelimi artırabilmektedir (Atalmış, 2022). Buna karşın çoktan seçmeli maddelerin üst düzey bilişsel süreçleri ortaya koymadaki sınırlılıkları, açık uçlu soruların ölçme ve değerlendirme süreçlerinde daha uygun bir seçenek olarak tercih edilmesine neden olmuştur. Günümüzde PISA ve TIMSS gibi uluslararası alanda yapılan pek çok sınavda çoktan seçmeli maddelere ek olarak öğrencilerin farklı zihinsel becerilerini ortaya koymalarını amaçlayan açık uçlu maddelere de yer verilmektedir (Aslanoglu, 2022). Çoktan seçmeli maddeler, beklenmeyen öğrenci yanıtlarını ortaya çıkarmak amacıyla açık uçlu maddeleri tamamlayıcı bir rol üstlenmektedir. Bir ölçme aracında açık uçlu ve çoktan seçmeli maddelerin birlikte kullanılmasının amacı, öngörülemeyen alt düzey davranış ve becerileri ortaya çıkarmaktır (Krosnick ve Presser, 2010). Bu bakımdan geliştirilen ölçme araçlarında her iki madde türünün bir arada yer alması sürdürülebilirlik gibi geniş bir tema çerçevesinde hem kapsam geçerliliğinin korunmasına hem de öğrencilerin üst düzey düşünme becerilerinin daha dengeli şekilde değerlendirilmesine olanak sağlayabilir.

Alan yazında sürdürülebilirlik konulu başarı testi çalışmalarının; çoğunlukla ortaokul öğrencileri (Arslan ve Görgülü-Arı, 2025; Burkaz-Ekinci, 2021; Somuncu-Demir, 2012; Şeker, 2017), lise öğrencileri (Belen, 2020) ile öğretmen adaylarına (Gökmen vd., 2017) yönelik geliştirildiği görülmektedir. Ayrıca, bu kapsamda geliştirilen testlerin tamamının çoktan seçmeli madde türünde yapılandırıldığı ve Bloom taksonomisinde daha alt düzey bilişsel becerilere yönelik olarak hazırlandığı belirlenmiştir. Sürdürülebilir çevre (Belen, 2020; Somuncu-Demir, 2012), sürdürülebilir kalkınma (Gökmen vd., 2017; Şeker, 2017) konularında geliştirilen çalışmaların yanı sıra doğrudan sürdürülebilirlik konusunda geliştirilen test çalışmalarının (Arslan ve Görgülü-Arı, 2025) alan yazında sınırlı olduğu saptanmıştır. Özellikle sürdürülebilirlik konusunun doğası gereği pek çok alt temayı içeren yapısı göz önüne alındığında özel yetenekli öğrencilerin bu alandaki bilgi ve becerilerinin farklı yollarla değerlendirilmesine duyulan ihtiyaç daha da önemli hale gelmektedir. Çoktan seçmeli ve açık uçlu maddelerin bir arada olduğu bir ölçme aracının hazırlanması öğrencilerin sürdürülebilirlik temasına ilişkin belirlenecek akademik yeterliklerin daha kapsamlı şekilde ortaya konulmasına olanak sağlayabilir. Bu gerekçelerden hareketle, ilkökul dördüncü ve ortaokul beşinci sınıf

düzeyindeki özel yetenekli öğrencilerin “Sürdürülebilirlik” ünitesine yönelik akademik başarılarını ölçmek amacıyla geçerli ve güvenilir bir ölçme aracı geliştirilmesi amaçlanmıştır. Bu amaç doğrultusunda aşağıdaki alt problemlere yanıt aranmıştır:

- (1) Özel yetenekli dördüncü ve beşinci sınıf öğrencileri için geliştirilen “Sürdürülebilirlik” başarı testinin geçerliğine ilişkin elde edilen sonuçlar nelerdir?
- (2) Özel yetenekli dördüncü ve beşinci sınıf öğrencileri için geliştirilen “Sürdürülebilirlik” başarı testinin güvenilirliğine ilişkin elde edilen sonuçlar nelerdir?

2. Yöntem

Araştırma kapsamında dördüncü ve beşinci sınıf düzeyinde öğrenim gören özel yetenekli öğrencilerin “Sürdürülebilirlik” ünitesi kapsamındaki başarısını ölçmek amacıyla test geliştirme sürecinde kullanılan araştırma modeli, evren ve örneklem, başarı testi geliştirme süreci, veri toplama araçları ve veri analizi bölümleri yer almaktadır.

2.1. Araştırmanın Modeli

Araştırma, Bilim ve Sanat Merkezleri’nde öğrenim gören dördüncü ve beşinci sınıf düzeyindeki özel yetenekli öğrenciler için “Sürdürülebilirlik” ünitesine yönelik bir başarı testi geliştirmeyi ve geliştirilen testin psikometrik özelliklerini incelemeyi amaçlayan test geliştirme ve psikometrik değerlendirme çalışmasıdır. Çalışma kapsamında test geliştirme sürecinin aşamaları yürütülmüş; ön uygulamadan elde edilen veriler üzerinden madde analizleri ile güvenilirlik ve geçerlik değerlendirmeleri gerçekleştirilmiştir. Psikometrik analizlerde kullanılan veriler, hedef gruptan ön deneme uygulaması kapsamında tek uygulamada toplanmıştır. Pilot uygulama ise test maddelerinin anlaşılabilirliğini ve uygulama sürecinin işleyişini değerlendirmek amacıyla ayrı bir öğrenci grubuyla gerçekleştirilmiştir.

2.2. Evren ve Örneklem

Araştırmanın evrenini Türkiye’de farklı bölgelerde yer alan Bilim ve Sanat Merkezleri’nde (BİLSEM) Destek Eğitim Programı (Destek-3) ve Bireysel Yetenekleri Fark Ettirme Programına (BYF-1) kayıtlı dördüncü ve beşinci sınıf öğrencileri oluşturmaktadır. Çalışmanın örneklemi, 2024–2025 eğitim öğretim yılının ikinci döneminde Türkiye’nin farklı bölgelerindeki Bilim ve Sanat Merkezlerinde (BİLSEM) öğrenim gören öğrenciler arasından ölçüt örnekleme yoluyla belirlenen ve etik kurul onayının ardından gerçekleştirilen ön deneme uygulamasına katılan toplam 238 öğrenci oluşturmıştır. Örnekleme dahil edilecek öğrencilerin BİLSEM’de öğrenim görüyor olmaları temel ölçüt olarak belirlenmiş; sınıf düzeyi açısından ise 4. sınıfta öğrenim gören öğrenciler ile 5. sınıfta olup ilgili dersi daha önce almış öğrenciler araştırma kapsamına dahil edilmiştir. 2024 Türkiye Yüzyılı Maarif Modeli kapsamında güncellenen BİLSEM Fen Bilimleri Dersi Çerçeve Öğretim Programı, Bilim ve Sanat Merkezlerinin yıllık planlamaları ve öğrenci profilinin özellikleri doğrultusunda içerik, süreç ve üründe farklılaştırmalara imkan tanıyan esnek bir çerçeve sunmaktadır. Tema ve kavramların ele alınma sırası ile derinliği BİLSEM’ler arasında değişkenlik gösterebilmekte ve buna bağlı olarak uygulama farklılıkları ortaya çıkabilmektedir. Bu çerçeve yapı dikkate alınarak geliştirilen testin farklı uygulama biçimlerinde kullanılabilirliğini ortaya koymak amacıyla çalışma grubu 4. ve 5. sınıf düzeylerini birlikte kapsayacak şekilde oluşturulmuştur. Örneklem, araştırmanın hedef kitlesi olan özel yetenekli öğrencileri temsil etmesi için BİLSEM’de öğrenim gören öğrencilerden seçilmiştir. Bunun nedeni, öğretim uygulamasının BİLSEM bağlamına göre geliştirilmiş olması ve uygulamanın BİLSEM’de yürütülmesidir. Beşinci sınıf öğrencileri için belirlenen “Sürdürülebilirlik ünitesini daha önce almış olma” ölçütü ise, öğrencilerin konuya ilişkin asgari ön bilgiye sahip olmasını sağlamayı hedeflemektedir.

Başarı testinin geçerlik ve güvenilirliğini değerlendirebilmek için ön deneme formunun yeterli sayıda öğrenciye uygulanmasına yönelik alan yazında araştırmacılar tarafından farklı görüşler yer almaktadır. Bu doğrultuda, Kline (1994) en az 200 kişilik bir örneklem grubunu önerirken; Tabachnick ve Fidell (2001), 150 ile 200 arasında bir örneklem büyüklüğünü yeterli görmektedir. Bu doğrultuda çalışmada ulaşılan 238 kişilik örneklem test maddelerinin analiz edilmesi için uygun bir büyüklük olarak kabul edilebilir. Ön deneme formuna katılan öğrencilere ilişkin bilgiler Tablo 1’de yer almaktadır.

Tablo 1. Sürdürülebilirlik Başarı Testi Ön Deneme Uygulamasına Katılan Öğrencilerin Demografik Özellikleri

Sınıf Düzeyi	Kız	Erkek	Toplam
4. Sınıf	39	58	97
5. Sınıf	62	79	141
Toplam	101	137	238

Tablo 1, ön deneme uygulamasına katılan öğrencilerin sınıf düzeyine ve cinsiyete göre dağılımını göstermektedir. Buna göre, 4. sınıftan 97; 5. sınıftan ise 141 öğrenci uygulamaya katılmıştır. Ön deneme uygulamasına katılan toplam 238 öğrencinin 101'inin kız, 137'sinin ise erkek olduğu görülmektedir.

2.3. Veri Toplama Araçları

Bu araştırmada, özel yetenekli öğrenciler için Fen Bilgisi dersi programındaki “Sürdürülebilirlik” ünitesi kapsamında düzey belirlemeye yönelik bir başarı testinin geliştirilmesi amaçlanmıştır. Sürdürülebilirlik ünitesine yönelik başarı testinin geliştirilmesi süreci; Crocker ve Algina (2006) tarafından ortaya konulan test geliştirme basamakları temel alınarak yapılandırılmıştır. Belirtilen basamaklara göre test geliştirme süreci; test puanlarının hangi amaçla kullanılacağına belirlenmesi, ölçülmek istenen yapıyı temsil eden davranışların tanımlanması, belirtke tablosunun hazırlanması, madde havuzunun oluşturulması, hazırlanan maddelerin uzman görüşüne sunulması, pilot uygulamanın gerçekleştirilmesi, testin daha geniş bir örneklemde ön deneme uygulamasının gerçekleştirilmesi, madde puanlarının istatistiksel özelliklerinin belirlenmesi, testin son formuna yönelik güvenilirlik ve geçerlik çalışmalarının yapılması ile testin puanlanması ve test sonuçlarının yorumlanmasına ilişkin yönergelerin geliştirilmesi aşamalarından meydana gelmektedir.

2.4. Test Amacının ve Kapsamının Belirlenmesi

Türkiye’de Bilim ve Sanat Merkezlerinde uygulanan programlar, akademik başarıyı ölçmekten ziyade, özel yetenekli öğrencilerin bilim, sanat ve teknoloji gibi alanlardaki güçlü yönlerini keşfetmelerine ve geliştirmelerine odaklanmaktadır. Bu kurumlarda öğrencilerin gelişimleri çoğunlukla öğretmen gözlemleri, ürün dosyaları ve benzeri süreç temelli değerlendirmelerle takip edilmekte; akademik başarıyı doğrudan ölçmeye yönelik standart testler kullanılmamaktadır (MEB, 2024). Bu nedenle, özel yetenekli öğrencilerin sürdürülebilirlik konusundaki yeterliklerini daha güvenilir şekilde değerlendirebilmek amacıyla bir başarı testi geliştirilmiştir. Geliştirilen testin amacı, BİLSEM’de 4. sınıf ve 5. sınıf düzeyinde öğrenim gören özel yetenekli öğrencilerin “Sürdürülebilirlik” ünitesine ilişkin akademik başarı düzeylerini programda hedeflenen öğrenme çıktılarıyla uyumlu şekilde geçerli ve güvenilir puanlarla belirlemektir. Testin kapsamı, MEB (2024) BİLSEM Fen Bilimleri dersi çerçeve öğretim programında “Sürdürülebilirlik” ünitesi kapsamında yer alan alt temalar, ana ve alt öğrenme çıktıları ile sınırlandırılmıştır. İçerik alanı, ünitenin temel kavramları ve alt boyutlarını temsil edecek şekilde belirtke tablosu üzerinden yapılandırılmıştır. Ölçme süreci kapsamında, öğrencilerin yalnızca hatırlama ve anlama basamaklarına yönelik düzeylerini değil, aynı zamanda uygulama–analiz–değerlendirme–yaratma gibi üst düzey bilişsel süreçleri ortaya koymalarını sağlayacak madde türleriyle akademik performanslarının çok yönlü şekilde değerlendirilmesi hedeflenmiştir.

2.5. Ölçülecek Özelliklerin Belirlenmesi ve Belirtke Tablosunun Oluşturulması

Bilim ve Sanat Merkezleri Fen Bilgisi dersi çerçeve programında yer alan sürdürülebilirlik konulu ünite incelendiğinde 10 ana tema altında 19 ana öğrenme çıktısı ve bu çıktılara bağlı toplam 66 alt öğrenme çıktısının yer aldığı belirlenmiştir. Geliştirilen testin Yenilenmiş Bloom Taksonomisi’nin (Krathwohl, 2002) üst düzey bilişsel süreç basamaklarını kapsayacak şekilde yapılandırılmasını sağlamak amacıyla, BİLSEM’lerde görev yapan farklı branşlardan öğretmenler ile akademisyenlerden oluşan bir uzman grubu (doktora mezunu sınıf öğretmeni, fen bilgisi öğretmeni, çift anadal mezunu öğretmen, iki program geliştirme uzmanı ve bir fen bilgisi alan uzmanı) öğretim programındaki öğrenme çıktılarının tamamını gözden geçirerek ilgili öğrenme çıktıları bilişsel düzey açısından yeniden değerlendirmiştir.

2.6. Madde Havuzunun Oluşturulması

Test maddeleri, Yenilenmiş Bloom Taksonomisi'nde yer alan bilişsel süreç becerileri temel alınarak hazırlanmıştır. Her bir ana öğrenme çıktısının ölçme aracında dengeli bir şekilde yer alması için her çıktıya en az üç madde yazılmasına dikkat edilmiş ve bu doğrultuda uzman görüşleri alınmadan önce 54 maddeden oluşan bir taslak madde havuzu hazırlanmıştır. Formda yer alan çoktan seçmeli maddeler dört seçenekten oluşmaktadır. Bu sayı, uzman görüşü öncesindeki ilk taslak havuzu ifade etmekte olup uzman değerlendirmeleri sonrasında oluşturulan ön deneme formunun madde sayısından farklıdır. Testte yer alan maddeler, “hatırlama, anlama, uygulama ve çözümleme” basamaklarına karşılık gelen öğrenme çıktılarında ağırlıklı olarak çoktan seçmeli sorularla yapılandırılmıştır. Değerlendirme basamağında hem çoktan seçmeli hem de açık uçlu sorulara yer verilirken yaratma düzeyine karşılık gelen öğrenme çıktıları açık uçlu sorularla değerlendirilmiştir.

2.7. Hazırlanan Maddelerin Uzman Görüşüne Sunulması

Ön deneme taslak formunda yer alan maddelerin Yenilenmiş Bloom Taksonomisi'ndeki bilişsel süreç becerilerine uygunluğunu değerlendirmek amacıyla farklı uzmanlık alanlarındaki araştırmacıların görüşlerine başvurulmuştur. Ön deneme taslak formunda yer alan maddelerin Yenilenmiş Bloom Taksonomisi'ndeki bilişsel süreç basamaklarına uygunluğunu değerlendirmek amacıyla uzman görüşüne başvurulmuştur. Maddelere ilişkin görüşlerine başvuru alan uzmanlar belirlenirken, tek tek uzmanların aynı ölçütleri karşılamasından ziyade uzman grubunun; özel yetenekliler alanı, fen bilimleri eğitimi (sürdürülebilirlik/çevre eğitimi), program geliştirme ve ölçme-değerlendirme alanlarında tamamlayıcı uzmanlıkları temsil edecek şekilde oluşturulmasına dikkat edilmiştir. Bu doğrultuda; özel yetenekliler alanında fen bilimleri eğitimi üzerine çalışan bir öğretim elemanı, program geliştirme alanında uzman iki akademisyen, bir ölçme ve değerlendirme uzmanı, BİLSEM'de görev yapan dört öğretmen (iki sınıf öğretmeni ve iki fen bilgisi/biyoloji öğretmeni) ile fen bilimleri eğitimi alanında sürdürülebilirlik üzerine çalışan bir öğretim elemanı olmak üzere toplam dokuz uzmandan görüş alınmıştır.

Uzmanlardan elde edilen geri bildirimler doğrultusunda her bir madde için Kapsam Geçerlik Oranı (KGO) hesaplanmıştır. Uzmanlara her madde için “Uygun”, “Düzeltilmeli” ve “Çıkarılabilir” seçenekleri sunulmuş; KGO değerleri Lawshe'nin (1975) formülüne göre belirlenmiştir. Dokuz uzman için Ayre ve Scally'nin (2014) önerdiği 0.78'lik ölçüt dikkate alınmış, bu değer altındaki maddeler içerik uygunluğu açısından yeniden değerlendirilmiştir. Buna göre, 0.78 ve üzerindeki maddeler kabul edilmiş; 0.67–0.75 aralığında kalan maddeler, uzman görüşleri doğrultusunda düzenlenerek tekrar değerlendirmeye sunulmuş; 0.44 ve altında kalan maddeler ise testten çıkarılmıştır.

KGO sonuçlarına göre, ön deneme formundaki 54 maddeden 38'i (%70) doğrudan kabul edilmiştir. 6'sı çoktan seçmeli, 3'ü açık uçlu olmak üzere toplam 9 maddede (%17) küçük düzenlemeler yapılarak maddeler testte tutulmuştur. 7 madde (%13) ise içerik uyumu yetersiz olduğu için testten çıkarılmıştır (5'i çoktan seçmeli, 2'si açık uçlu). Testin toplam kapsam geçerlik indeksi 0.85 olarak hesaplanmış ve bu değer, kabul edilebilir eşik değer üzerinde olduğundan testin kapsam geçerliliğinin istatistiksel olarak yeterli olduğu sonucuna ulaşılmıştır. Uzman görüşleri sonrasında oluşturulan ön deneme formunda 38 dört seçenekli çoktan seçmeli madde ile dokuz açık uçlu madde yer almaktadır. Böylece form, soru düzeyinde toplam 47 maddeden oluşmaktadır. İlk açık uçlu madde dört ayrı alt soruya ayrıldığından, ön deneme formunda toplam 50 puanlanabilir yanıt birimi bulunmaktadır.

2.8. Pilot Uygulamanın Gerçekleştirilmesi

Oluşturulan taslak test formu, uygulama sürecinin işleyişini ve maddelerin anlaşılabilirliğini değerlendirmek amacıyla Bilim ve Sanat Merkezleri'nde 4. ve 5. sınıf düzeyinde öğrenim gören 5 öğrenciye uygulanmıştır. Bu uygulama sırasında hem öğrencilerin testle etkileşimleri yakından gözlenmiş hem de öğrencilerden kısa geri bildirimler alınmıştır. Süreçte yönergelerin ne kadar açık olduğu, öğrencilerin madde köklerini ve seçenekleri nasıl anladığı, kullanılan kavram ve terimlerin sınıf düzeyine uygunluğu, görsellerin ve şekillerin okunabilirliği ile maddeyi destekleme düzeyi, öğrencilerin zorlandığı noktalar ve uygulamanın genel akışı ile süresi değerlendirilmiştir.

Pilot uygulamadan elde edilen değerlendirmeler doğrultusunda taslak form üzerinde çeşitli düzenlemeler yapılmıştır. Bu kapsamda toplam 11 maddede dil ve ifade sadeleştirilmiş, bazı cümleler kısaltılmış ve anlam belirsizliği oluşturabilecek kullanımlar düzeltilmiştir. Ayrıca, 4 maddede kavram karmaşasına yol açabilecek ifadeler

netleştirilerek gerekli görülen yerlerde kısa açıklayıcı ipuçları eklenmiştir. Görsel ve şekil kullanılan maddelerde ise 5 madde için görsellerin boyutu, çözünürlüğü ve etiketleri yeniden düzenlenmiştir. İki maddede görsel ve şeklin maddeyle ilişkisini güçlendirmek amacıyla soru içerisinde yeniden konumlandırılmaları yapılmıştır. Bunun yanında yönergeler daha anlaşılır hale getirilmiş; sayfa düzeni, boşluklar ve madde yerleşimi öğrencilerin okuma akışını destekleyecek şekilde güncellenmiştir. Pilot uygulamada gözlenen süre ve akış dikkate alınarak testin madde sırası korunmuştur. Bu düzenlemeler sonucunda form, ön deneme uygulamasına hazır hale getirilmiştir.

2.9. Ön Deneme Formunun Uygulanması ve Puanlanması

Gerekli etik ve kurumsal izinlerin alınmasının ardından gerçekleştirilen pilot uygulamadan elde edilen geri bildirimler doğrultusunda son şekli verilen ön deneme formu, Türkiye'nin farklı bölgelerindeki Bilim ve Sanat Merkezlerinde öğrenim gören 4. ve 5. sınıf düzeyindeki toplam 238 öğrenciye uygulanmıştır. Bu aşama, testin psikometrik özelliklerini incelemeye yönelik analizler için veri elde etmeyi amaçlamıştır. Uygulama sürecinde, testin güvenilirliğini olumsuz etkileyebilecek durumların en aza indirilmesi amacıyla gerekli önlemler alınmış; uygulama ortamı, yönergelerin açıklığı ve öğrencilerin dikkat düzeyleri hem öğretmenler hem de araştırmacı tarafından izlenmiştir.

Ön deneme formunda yer alan 38 çoktan seçmeli maddenin her biri 1 puan üzerinden değerlendirilmiştir. Dokuz açık uçlu maddenin ilki dört alt sorudan oluştuğu için açık uçlu bölümde toplam 12 puanlanabilir yanıt birimi bulunmaktadır. Açık uçlu yanıtların her biri 2-1-0 puanlama anahtarıyla değerlendirilmiş; buna göre ön deneme formundan alınabilecek en yüksek puan çoktan seçmeli bölüm için 38, açık uçlu bölüm için 24 ve testin geneli için 62 olarak belirlenmiştir.

2.10. Verilerin Analizi

Testin ön deneme formundan elde edilen veriler çoktan seçmeli ve açık uçlu maddelerin farklı yapılarından dolayı iki ayrı analiz sürecine tabi tutulmuştur.

Çoktan seçmeli maddeler için öğrencilerin her bir maddeye ilişkin verdikleri yanıtlar Excel formatına aktararak doğru cevaplar için "1" puan, yanlış cevaplar için "0" şeklinde kodlanmıştır. Kodlanan veriler toplam puana dönüştürülerek düşükten yükseğe doğru sıralanmış, ardından alt ve üst grupların belirlenmesi amacıyla en düşük ve en yüksek %27'lik dilimler oluşturulmuştur. Çoktan seçmeli maddelere ilişkin; güçlük indeksleri (p), madde ayırt edicilik indeksleri (r), test geneline ilişkin ayırt edicilik ve güçlük değerleri, test ortalaması, standart sapma ile güvenilirlik katsayıları (KR20 ve KR21) TAP (Test Analysis Program Version 19.1.4) aracılığıyla analiz edilmiştir.

Açık uçlu maddelerde ise süreç çoktan seçmeli maddelere göre farklı şekilde yürütülmüştür. Açık uçlu maddeler için puanlama araştırmacı tarafından hazırlanan rubrik doğrultusunda 2-1-0 şeklinde yapılmıştır. Öğrencilerin verdikleri cevaplar öncelikle araştırmacı tarafından hazırlanan rubrik doğrultusunda iki bağımsız puanlayıcı tarafından değerlendirilmiştir. Puanlayıcılar tarafından verilen puanlar Excel ortamına aktararak her bir madde için puan dağılımları oluşturulmuştur. Açık uçlu maddeler için puanlama, araştırmacı tarafından hazırlanan rubrik doğrultusunda 2-1-0 puanlama anahtarı kullanılarak iki bağımsız puanlayıcı tarafından gerçekleştirilmiştir. Puanlayıcıların verdiği puanların ortalaması alınarak her bir yanıt birimi için madde güçlüğü ve madde ayırt edicilik indeksleri hesaplanmıştır. Ayrıca, puanlama sürecindeki tutarlılığı belirlemek amacıyla puanlayıcılar arası uyum Intraclass Correlation Coefficient (ICC) ile incelenmiştir.

3. Bulgular

Araştırmanın bu bölümünde özel yetenekli dördüncü ve beşinci sınıf düzeyindeki öğrencilere yönelik geliştirilen "Sürdürülebilirlik" ünitesine yönelik akademik başarı testinin güvenilirlik, geçerlik ve madde analizine ilişkin bulgular yer almaktadır.

3.1. Sürdürülebilirlik Akademik Başarı Testinin Geçerliğine Yönelik Bulgular

2024 Maarif Modeli kapsamında Fen Bilimleri dersi öğretim programında yer alan "Sürdürülebilirlik" ünitesi kapsamında geliştirilen akademik başarı testine ait hazırlanan belirtke tablosu Tablo 2'de yer almaktadır.

Tablo 2. Uzman Görüşü Öncesi Taslak Sürdürülebilirlik Akademik Başarı Testine Ait Belirtke Tablosu

Bilgi Aşaması	Bilişsel Süreç Becerileri					
	Hatırlama	Anlama	Uygulama	Çözümleme	Değerlendirme	Yaratma
Olgusal	FEN.DEP. 1.1.a FEN.DEP. 3.1.a FEN.DEP. 3.1.b			FEN.DEP.8.1. b		
Kavramsal	FEN.DEP. 3.2.b	FEN.DEP.1.1. b FEN.DEP.1.2. a D.E.K.3 FEN.DEP.9.2. a D.E.K.8	FEN.DEP.6.2 .c	FEN.DEP.1.1. d D.E.K.1 D.E.K.2 FEN.DEP.1.2. b FEN.DEP.1.2. c FEN.DEP.1.2. d FEN.DEP.2.1. a FEN.DEP.2.2. b FEN.DEP.4.1. a FEN.DEP.5.2. ç FEN.DEP.6.1. ç FEN.DEP.6.2. b	D.E.K.4 D.E.K.5 FEN.DEP.2.2.a FEN.DEP.4.1.b FEN.DEP.4.1.d FEN.DEP.4.2.b FEN.DEP.4.2.d D.E.K.6 FEN.DEP.8.1.c FEN.DEP.9.1.c FEN.DEP.9.2.b FEN.DEP.10.1.c D.E.K.7	FEN.DEP.1. 2.ç FEN.DEP.2. 1.b FEN.DEP.3. 2.c FEN.DEP.5. 2.a FEN.DEP.4. 1.c FEN.DEP.6. 1.a FEN.DEP.6. 1.d
İşlemsel			FEN.DEP.5.1 .a FEN.DEP.6.1 .c			FEN.DEP.3. 2.a FEN.DEP.4. 2.a FEN.DEP.10 .1.a
Biliş ötesi			FEN.DEP.7.1 .c	FEN.DEP.7.2. ç	FEN.DEP.7.2.d FEN.DEP.10.2.a FEN.DEP.10.2.b D.E.K.9	
Taslak Madde Havuzundaki Madde Sayısı	4	5	4	14	17	10

Not: D.E.P. kodlu öğrenme çıktıları Bilim ve Sanat Merkezi Fen Bilimleri dersi çerçeve programında yer almaktadır. D.E.K. kodlu öğrenme çıktıları dışarıdan eklenen kazanımlar olarak yer almaktadır.

Tablo 2’de sunulan öğrenme çıktılarının kodlama yapısını daha anlaşılır şekilde göstermek amacıyla, çerçeve programda yer alan FEN.DEP kodlu bir öğrenme çıktısı ile araştırmacı tarafından eklenen D.E.K kodlu bir öğrenme çıktısı üzerinden örnekleme yapılmıştır. Çerçeve program kapsamında FEN.DEP.1.1. “Sürdürülebilirlik kavramının gündelik hayata etkilerini sorgulayabilme” ana öğrenme çıktısı yer almakta; bu ana çıktı altında a) “Sürdürülebilirlik kavramını tanımlar” alt öğrenme çıktısı bulunmaktadır. Buna ek olarak, D.E.K.1. “Sürdürülebilirliğin boyutlarını ayırıştırır” ifadesi ise çerçeve programda yer almayan ve araştırmacı tarafından programa eklenen ana öğrenme çıktısını temsil etmektedir.

Sürdürülebilirlik konulu başarı testinin geliştirilmesi sürecinde, uzman görüşü öncesindeki taslak madde havuzu için hazırlanan belirtke tablosunda 4 madde hatırlama, 5 madde anlama, 4 madde uygulama, 14 madde çözümleme, 17 madde değerlendirme ve 10 madde yaratma düzeyine yönelik olmak üzere toplam 54 madde yer almaktadır.

3.2. Sürdürülebilirlik Akademik Başarı Testinin Madde Analizine Yönelik Bulgular

3.2.1. Çoktan Seçmeli Maddelere İlişkin Bulgular

Çalışma kapsamında geliştirilmesi amaçlanan başarı testinin çoktan seçmeli maddeler için üst ve alt grup öğrenci sayıları, madde ayırt edicilik indeksleri, madde güçlük indeksleri ve güvenilirlik hesaplamaları Test Analysis Program (TAP) ile yapılmıştır. Ön deneme formunda yer alan 38 dört seçenekli çoktan seçmeli maddenin puanlamasında her doğru cevap için 1 puan, yanlış ya da boş bırakılan yanıt için 0 puan kullanılmıştır.

Akademik başarı testinin ön deneme uygulaması 238 özel yetenekli 4. ve 5. sınıf öğrencisiyle gerçekleştirildikten sonra, öğrencilerin test puanları yüksekte düşüğe doğru sıralanarak madde analizi için üst ve alt gruplar şeklinde gruplandırılmıştır. Geleneksel madde analizi yaklaşımına göre sıralanan puanların yaklaşık %27'si üst grubu, son %27'si ise alt grubu oluşturmaktadır. Bu oran temel alındığında, çalışmada üst grup 64 öğrenci, alt grup ise 64 öğrenci olarak belirlenmiştir. İlk madde analizi sonucunda, çoktan seçmeli maddelerden güçlük değeri 0.30–0.80 aralığının dışında kalan 2., 4. ve 6. maddeler ile ayırt edicilik katsayısı 0.25'in altında olduğu belirlenen 5., 13., 18., 23., 32. ve 38. maddeler testten çıkarılmıştır. Bu ilk değerlendirmeye göre testin ortalama madde güçlüğü 0.51, ortalama madde ayırt ediciliği 0.38 olarak bulunmuştur. Güvenirliliğe ilişkin katsayılar ise KR-20 için 0.79, KR-21 için 0.77 olarak hesaplanmış ve testin ilk formunun kabul edilebilir düzeyde güvenilirlik sunduğu görülmüştür. İlgili maddelerden çıkartıldıktan sonra yapılan ikinci madde analizi sonucu Tablo 3'te gösterilmiştir.

Tablo 3. Madde Analizi Sonucunda Güncellenen Başarı Testinin Çoktan Seçmeli Maddelerin Güçlük ve Ayırt Edicilik İndeksleri

Madde No	Güçlük İndeksi (p)	Ayırt Edicilik İndeksi (r)	Madde No	Güçlük İndeksi (p)	Ayırt Edicilik İndeksi (r)
1	.70	.39	22	.52	.61
3	.61	.37	24	.60	.63
7	.71	.32	25	.60	.56
8	.65	.46	26	.46	.40
9	.45	.36	27	.58	.65
10	.65	.55	28	.68	.44
11	.51	.38	29	.61	.53
12	.70	.45	30	.40	.49
14	.49	.30	31	.50	.58
15	.35	.37	33	.37	.43
16	.52	.42	34	.72	.61
17	.34	.27	35	.41	.52
19	.66	.72	36	.42	.30
20	.53	.52	37	.52	.42
21	.53	.51			

Başarı testinin ön deneme formunun uygulanmasının ardından yapılan analiz sonucunda çoktan seçmeli toplam 9 madde testten çıkarılmış ve kalan 29 madde üzerinden madde analizi yeniden gerçekleştirilmiştir. Yapılan ikinci analizde, testin ortalama madde güçlük indeksi 0.55, ortalama madde ayırt edicilik indeksi 0.47 olarak hesaplanmıştır. Testin güvenilirliğine ilişkin katsayılar incelendiğinde ise KR-20 güvenilirlik katsayısının 0.81, KR-21 katsayısının ise 0.80 düzeyinde olduğu belirlenmiştir. Çoktan seçmeli testlere ilişkin literatürde farklı öneriler bulunmakla birlikte, madde güçlük indeksinin 0.30 ile 0.80 aralığında yer alması genellikle orta düzey güçlüğü ortaya koyduğu; madde ayırt edicilik indeksinin 0.25'in üzerinde olmasının ise maddenin öğrencileri başarı düzeylerine göre yeterince ayırt ettiğini gösterdiği belirtilmektedir (Uddin vd., 2020). Instructional Assessment Resources (IAR, 2011) ise madde

güçlüğünü; 0.90 ve üzerini “kolay”, 0.20–0.90 aralığını “orta”, 0.20’nin altını ise “zor” olarak sınıflandırmaktadır. Bu ölçütlerle birlikte değerlendirildiğinde madde analizi sonucunda; testin hedef kitle için uygun güçlükte olduğu, maddelerin büyük bölümünün ayırt edicilik açısından yeterli değerlere sahip olduğu ve testin güvenilirlik katsayısının kabul edilebilir aralıkta yer aldığı belirlenmiştir.

3.2.2. Açık Uçlu Maddelere İlişkin Bulgular

Ön deneme formunda yer alan açık uçlu maddeler “Sürdürülebilirlik” ünitesi kapsamındaki ilgili öğrenme çıktılarıyla ilişkilendirilerek hazırlanmıştır. Ön deneme formunda yer alan açık uçlu maddelerin öğrenme çıktılarıyla eşleştirilmesine ilişkin bilgiler Tablo 4’te yer almaktadır.

Tablo 1. Ön Deneme Formunda Yer Alan Açık Uçlu Maddelerin Öğrenme Çıktılarıyla Eşleştirilmesi

Madde No	İlişkilendirilen Öğrenme Çıktısı	Bilgi Boyutu	Bilişsel Süreç Becerisi
Soru 1-a	FEN.DEP.1.2.ç	Kavramsal	Yaratma
Soru 1-b	FEN.DEP.1.2.ç	Kavramsal	Yaratma
Soru 1-c	FEN.DEP.1.2.ç	Kavramsal	Yaratma
Soru 1-d	FEN.DEP.1.2.ç	Kavramsal	Yaratma
Soru 2	FEN.DEP.2.1.b	Kavramsal	Yaratma
Soru 3	FEN.DEP.3.2.a	İşlemsel	Yaratma
Soru 4	FEN.DEP.3.2.c	Kavramsal	Yaratma
Soru 5	FEN.DEP.4.2.a	İşlemsel	Yaratma
Soru 6	FEN.DEP.4.2.b	Kavramsal	Değerlendirme
Soru 7	FEN.DEP.5.2.a	Kavramsal	Yaratma
Soru 8	FEN.DEP.6.1.a	Kavramsal	Yaratma
Soru 9	FEN.DEP.6.1.d	Kavramsal	Yaratma

Not: Bilgi boyutu ve bilişsel süreç sınıflaması Tablo 2’deki belirtke tablosu temel alınarak belirlenmiştir.

Ön deneme formunda dokuz açık uçlu soru yer almakta; ilk maddenin dört ayrı alt sorudan oluşması nedeniyle açık uçlu bölüm toplam 12 puanlanabilir yanıt biriminden oluşmaktadır. Bu yanıt birimleri araştırmacı tarafından hazırlanan rubrik doğrultusunda 2-1-0 puanlama anahtarı kullanılarak değerlendirilmiştir. Ön deneme formunda yer alan açık uçlu maddelerin analizleri ve bu maddelere ilişkin güvenilirlik hesaplamaları, Klasik Test Kuramı çerçevesinde yürütülmüştür. Bu doğrultuda, her bir madde için öğrencilerin aldıkları puanlar dikkate alınarak madde güçlüğü (p) ve madde ayırt ediciliği (D) değerleri hesaplanmıştır (Şata ve Karakaya, 2022). Buna göre, madde güçlüğü (p), öğrencilerin ilgili maddeden aldıkları ortalama puandan, o madde için verilebilecek en düşük puanın çıkarılması ve elde edilen sonucun maddenin puan aralığını oluşturan en yüksek puan ile en düşük puan arasındaki farka bölünmesiyle hesaplanmaktadır. Başarı testinde yer alan açık uçlu maddelere ilişkin madde ortalamaları ile madde güçlük indekslerine ilişkin bilgiler Tablo 5’te gösterilmiştir.

Tablo 5. Sürdürülebilirlik Başarı Testindeki Açık Uçlu Maddelerin Ortalamaları ve Madde Güçlük İndeksleri

Madde No	Puanlayıcı 1		Puanlayıcı 2		Puanlayıcı Ortalaması	
	Madde Ort.	Madde Güç.İnd.	Madde Ort.	Madde Güç.İnd.	Madde Ort.	Madde Güç.İnd.
Soru 1-a	0.88	0.44	0.88	0.44	0.88	0.44
Soru 1-b	0.85	0.43	0.76	0.38	0.81	0.41

Soru 1-c	0.48	0.24	0.39	0.20	0.44	0.22
Soru 1-d	0.97	0.49	0.88	0.44	0.93	0.47
Soru 2	1.21	0.61	1.15	0.58	1.18	0.60
Soru 3	0.97	0.49	0.94	0.47	0.96	0.48
Soru 4	0.81	0.41	0.81	0.41	0.81	0.41
Soru 5	1.24	0.62	1.24	0.62	1.24	0.62
Soru 6	0.88	0.44	0.88	0.44	0.88	0.44
Soru 7	1	0.50	1	0.50	1	0.50
Soru 8	1.06	0.53	1	0.50	1.03	0.52
Soru 9	0.88	0.44	0.85	0.43	0.87	0.44

Açık uçlu maddelere ilişkin puanlayıcı ortalamalarına dayalı olarak hesaplanan madde güçlük indeksleri (Tablo 5) incelendiğinde, 12 puanlanabilir yanıt biriminin 10'unun 0.40 ile 0.60 aralığında yer aldığı görülmektedir. Bu aralığın dışında kalan yanıt birimlerinden 1-c'nin güçlük indeksinin 0.22, 5. sorunun güçlük indeksinin ise 0.62 olduğu belirlenmiştir. Instructional Assessment Resources (IAR, 2011) madde güçlük düzeylerini kolay (0.90 ve üzeri), orta (0.20–0.90) ve zor (0.20'nin altı) şeklinde sınıflandırmaktadır. Bu ölçüt temel alındığında, çalışmadaki açık uçlu maddelerin genel olarak orta güçlük düzeyine sahip olduğu söylenebilir. Açık uçlu soruların madde ayırt ediciliği (D) analiz edilirken, öğrencilerin toplam test puanları temel alınarak en başarılı %27'lik dilim üst grup, en düşük başarı düzeyindeki %27'lik dilim ise alt grup olarak belirlenmiştir. İki grubun maddeye ilişkin ortalama puanları karşılaştırılarak, aralarındaki fark maddenin alabileceği puan aralığına göre standartlaştırılmıştır. Bu işlem sonucunda elde edilen değer, ilgili maddenin yüksek başarı düzeyindeki öğrencilerle düşük başarı düzeyindeki öğrencileri ne ölçüde ayırt edebildiğini göstermektedir. Bu kapsamda açık uçlu maddelerin ayırt edicilik analizinde, iki bağımsız uzman tarafından verilen puanlar ilk olarak ayrı ayrı incelenmiş, ardından her madde için bu puanların ortalaması alınarak analizde kullanılacak nihai puan seti oluşturulmuştur. Bu ortalama puanlar üzerinden maddelerin ayırt edicilik indeksleri hesaplanmıştır. Açık uçlu maddelere ilişkin elde edilen ayırt edicilik değerleri ve her bir maddenin performansına yönelik ayrıntılı bilgiler Tablo 6'da sunulmaktadır.

Tablo 6. Sürdürülebilirlik Başarı Testindeki Açık Uçlu Maddelerin Ayırt Edicilik İndeksleri

Madde No	Puanlayıcı 1		Puanlayıcı 2		Madde Ayırt Edicilik İnd. (D)	Ortalama Madde Ayırt Edicilik İnd.
	Üst Grubun Madde Ort.	Alt Grubun Madde Ort.	Üst Grubun Madde Ort.	Alt Grubun Madde Ort.		
S1-a	1.33	0.67	0.33	1.40	0.60	0.37
S1-b	1.47	0.60	0.43	1.40	0.73	0.38
S1-c	1.27	0.60	0.33	1.27	0.53	0.35
S1-d	1.53	0.73	0.40	1.47	0.80	0.37
S2	1.73	0.87	0.43	1.47	0.80	0.38
S3	1.67	0.47	0.60	1.47	0.33	0.59
S4	1.20	0.40	0.40	1.07	0.40	0.37
S5	1.67	0.40	0.64	1.53	0.40	0.60
S6	1.60	0.13	0.74	1.33	0.13	0.67
S7	1.53	0.13	0.70	1.27	0.13	0.63
S8	1.67	0.20	0.73	1.60	0.20	0.72
S9	1.66	0.60	0.53	1.53	0.60	0.50

Açık uçlu maddeler için puanlayıcı ortalamalarına dayalı olarak yapılan analiz sonucunda, maddelerin ayırt edicilik indekslerinin 0.35 ile 0.72 aralığında değiştiği belirlenmiştir (Tablo 6). Instructional Assessment Resources (IAR, 2011), ayırt edicilik değerlerini 0.30–0.39 aralığında “iyi”, 0.40 ve üzerindeki değerleri ise “çok iyi” düzeyde ayırt

edicilik olarak sınıflandırmaktadır. Bu ölçütler dikkate alındığında, çalışma kapsamında yer alan tüm açık uçlu maddelerin iyi ya da çok iyi düzeyde ayırt edicilik gösterdiği belirlenmiştir. Başarı testindeki açık uçlu maddelerin güvenilirliği, Klasik Test Kuramı çerçevesinde puanlayıcılar arası tutarlılığa dayalı olarak belirlenmiştir. Bu amaçla, iki bağımsız puanlayıcının verdiği puanlar arasındaki uyumu inceleyen sınıf içi korelasyon katsayısı (ICC) hesaplanmıştır. Alan yazında, ICC değerleri için 0.40 düzeyi kabul edilebilir alt sınır olarak önerilmektedir (Şata ve Karakaya, 2022). Çalışmada elde edilen madde bazlı ICC değerleri Tablo 7'de sunulmakta olup bu değerler açık uçlu maddelerin puanlanmasında puanlayıcılar arasında yüksek düzeyde tutarlılık sağlandığını göstermektedir.

Tablo 7. Açık Uçlu Maddelerin Klasik Test Kuramına Göre Kestirilen Puanlayıcı Güvenirlik Değerleri

	Sınıf içi korelasyon katsayısı	%95 güven aralığı	P
Maddeler bazında	0.979	0.917–0.992	0.000
Toplam puan bazında	0.989	0.957–0.996	0.000

Tablo 7'de belirtildiği üzere, açık uçlu maddelerin, maddeler bazında elde edilen sınıf içi korelasyon değeri (ICC) 0.979 olarak saptanmıştır. Bu sonuç, puanlayıcıların maddeleri ayrı ayrı değerlendirirken tutarlı ve benzer puanlar verdiklerini göstermektedir. Aynı maddelerin toplam puanları üzerinden hesaplanan ICC değerinin ise 0.989 olduğu belirlenmiştir. ICC, 0 ile 1 arasında bir değerdir; 0.5'in altındaki değerler zayıf güvenilirliği, 0.5 ile 0.75 arasındaki değerler orta düzeyde güvenilirliği, 0.75 ile 0.90 arasındaki değerler iyi güvenilirliği ve 0.90'un üzerindeki bir değer mükemmel güvenilirliği göstermektedir (Koo ve Li, 2016). Elde edilen bu sonuç, iki puanlayıcının verdiği puanların ortalamasına dayalı toplam puanın yüksek bir güvenilirliğe sahip olduğunu ortaya koymaktadır. Ek olarak analizlerde elde edilen tüm sınıf içi korelasyon değerlerinin istatistiksel olarak anlamlı olduğu görülmektedir ($p=0.000$).

4. Sonuç, Tartışma ve Yorum

Çalışma kapsamında 2024 yılında yayımlanan Maarif Modeli Bilim ve Sanat Merkezleri Fen Bilimleri dersi çerçeve öğretim programında bulunan öğrenme çıktıları temel alınarak bir başarı testi geliştirilmiştir. Geliştirilen test, sürdürülebilirlik konusuna ilişkin bilgi ve becerileri ölçmeye yönelik olup dört seçenekli 29 çoktan seçmeli madde ile dokuz açık uçlu sorudan oluşmaktadır. İlk açık uçlu soru dört ayrı alt soruya ayrıldığından, açık uçlu bölümde 12 puanlanabilir yanıt birimi bulunmaktadır. Buna göre testte soru düzeyinde 38 madde, puanlama düzeyinde ise 41 yanıt birimi yer almaktadır. Çoktan seçmeli maddelerin her doğru cevabı 1 puan; açık uçlu yanıt birimlerinin her biri ise doğru yanıt için 2 puan, kısmen doğru yanıt için 1 puan ve yanlış ya da boş yanıt için 0 puan üzerinden değerlendirilmektedir. Testten alınabilecek en yüksek puan çoktan seçmeli bölüm için 29, açık uçlu bölüm için 24 ve testin geneli için 53'tür.

Sürdürülebilirlik başarı testindeki çoktan seçmeli maddeler için gerçekleştirilen analizde, testin ortalama madde güçlük indeksi 0.55, ortalama madde ayırt edicilik indeksi 0.47 olarak elde edilmiştir. Testin güvenilirliğine ilişkin katsayılar incelendiğinde ise KR-20 güvenilirlik katsayısının 0.81, KR-21 katsayısının ise 0.80 düzeyinde olduğu belirlenmiştir. Bu değerler, testin genel olarak orta düzeyde güçlüğüne sahip, iyi ayırt ediciliğe ve yüksek güvenilirlik düzeyine sahip bir ölçme aracı olduğunu göstermektedir. Açık uçlu maddelere ilişkin puanlayıcı ortalamalarına dayalı olarak hesaplanan madde güçlük indekslerinin 0.22 ile 0.62; farklı puanlayıcıların verdiği puanların ortalaması alınarak hesaplanan madde ayırt edicilik indekslerinin ise 0.35 ile 0.72 arasında değiştiği saptanmıştır. Elde edilen bu sonuçlara göre açık uçlu maddelerin genel olarak orta güçlük düzeyine sahip olduğu, madde ayırt ediciliği olarak iyi ya da çok iyi düzeyde ayırt edicilik gösterdiği belirlenmiştir. Alan yazında ortalama madde güçlüğü 0.40 ile 0.60 arasında, madde ayırt ediciliğinin 0.30 ve üzerinin ayırt edici olduğu ve testte kalabileceği, 0.40'ın üzerindeki değerlerin ise mükemmel ayırt ediciliğe sahip olduğu bildirilmiştir (İlhan, 2022). Özel yetenekli öğrencilere yönelik geliştirilen testlerde sınıf düzeyinde geliştirilen testlerin kolay seviyede kalması puanların dar bir aralıkta toplanmasına yol açabilmekte ve bu durum testin ayırt ediciliğini sınırlayabilmektedir. Bu nedenle, testin daha yüksek tavan aralığı ve geniş puan dağılımı üretecek şekilde yapılandırılmış olması ölçme sürecinin ayırt ediciliğini destekleyebilir (Warne, 2012). Bu açıdan değerlendirildiğinde, açık uçlu maddelerdeki güçlük ve ayırt edicilik dağılımı, testin öğrenciler arasındaki başarı düzeyi farklarını daha net şekilde ortaya koyabildiği yorumunu desteklemektedir.

Testte yer alan açık uçlu maddelerin toplam puanları üzerinden hesaplanan ICC değerinin ise 0.989 olduğu belirlenmiştir. Bu bulgu, açık uçlu maddelerin puanlanmasında puanlayıcılar arası tutarlılığın çok yüksek olduğunu

ve puanlamaya bağlı ölçme hatasının sınırlı kaldığını göstermektedir. Benzer şekilde, ortaokul öğrencilerine yönelik sürdürülebilirlik konusunda başarı testi geliştiren Arslan ve Görgülü-Arı (2025) testin ortalama güçlük değerinin 0.52, ortalama ayırt edicilik değerinin ise 0.59 olduğunu belirlemiştir. Ayrıca, testin KR-20 güvenirlik katsayısı 0.84; Cronbach Alpha katsayısı ise 0.87 olarak elde edilmiştir. Öğrenme konusunda yapılan çalışmalarda öğrenmenin orta zorluk düzeyinde işlevsel olduğu öne çıkmaktadır (Şahin, 2018). Bu bakımdan özel yetenekliler için oluşturulan program hedeflerinde bu durum göz ardı edilmemelidir (Emir, 2021). Dolayısıyla, program hedefleri ile ölçme ve değerlendirme süreçleri arasındaki uyum için hazırlanan test maddelerinin uygun güçlük ve ayırt edicilik düzeylerinde olması önemlidir. Yapılan bu çalışmada geçerlik, test puanlarının amaçlanan yorum ve kullanımına yönelik kanıtların değerlendirilmesi çerçevesinde tartışılmıştır. Nitekim eğitsel ve psikolojik testler için belirlenen standartlarda, geçerlik ifadelerinin belirli yorum ve kullanım amaçlarına dayanması gerektiğini ve “testin geçerliği” gibi genelleşici bir ifadenin uygun olmadığını vurgulamaktadır (AERA, APA, & NCME, 2014). Bu çalışmada test puanlarının amaçlanan kullanımı, BİLSEM Fen Bilimleri çerçeve programında “Sürdürülebilirlik” ünitesi kapsamında tanımlanan öğrenme çıktıları doğrultusunda öğrencilerin akademik başarısını değerlendirmek olarak belirlenmiştir. Sireci ve Faulkner-Bond (2014) ile Kane (2013), bu tür bir kullanımda puan yorumlarının dayandığı varsayımların açıkça ortaya konulmasının ve sunulan kanıtlarla desteklenmesinin kritik olduğunu vurgulamaktadır. Bu nedenle elde edilen bulgular, testin amaçlanan kullanımını desteklemekle birlikte, bu çalışmada sunulan kanıtların kapsamı dışında kalan farklı kullanım amaçları için ayrıca ele alınmalıdır.

Özel yetenekli öğrencilerin eğitiminde Bloom taksonomisindeki hatırlama ve anlama düzeylerindeki hedeflere daha az, uygulama basamağı ve üzerindeki bilişsel düzeylere daha fazla ağırlık verilmesi gerektiği öne çıkmaktadır (Emir, 2021). Bu yaklaşım, başarı testinde hedeflenen bilişsel düzeylerin açık şekilde tanımlanmasını ve testin madde örneklerinin üst düzey bilişsel süreç basamaklarını temsil edecek şekilde yapılandırılmasını öne çıkarmaktadır. Bilim ve Sanat Merkezleri Fen Bilimleri dersi çerçeve öğretim programındaki öğrenme çıktıları uzman görüşlerine dayalı olarak yeniden değerlendirilerek bazı öğrenme çıktıları güncellenmiş, bazıları ise programa eklenmiştir. Geliştirilen başarı testinde Bloom Taksonomisinin üst düzey bilişsel süreçlerini ölçebilecek maddelere ağırlık verilmesi amaçlanmıştır. Bu bağlamda, özel yetenekli öğrencilerin sürdürülebilirlik konusundaki akademik yeterliklerini daha kapsamlı şekilde değerlendirmek için testte hem çoktan seçmeli hem de açık uçlu maddelere birlikte yer verilmiştir. Her iki madde türünün avantajları ve sınırlılıkları birlikte değerlendirildiğinde geliştirilen testin birbirini tamamlayan özelliklere sahip bütüncül bir ölçme yapısı sunduğu söylenebilir. Öyle ki; çoktan seçmeli maddeler öğrencilerin konuya yönelik temel kavrayışı nesnel şekilde ölçmeye ve değerlendirmeye olanak sağlarken, açık uçlu maddeler öğrencilerin akıl yürütme, ilişkilendirme ve yorumlama becerilerini ortaya koymasına imkan tanıyabilir. Bu açıdan alt ve üst düzey bilişsel süreçleri kapsayan akademik başarı testlerinde her iki soru tipi birlikte kullanılabilir (Kızılkaya ve Seven, 2017). Alan yazında da sürdürülebilirlik konusunda geliştirilen benzer başarı testlerinin çoğunlukla sadece çoktan seçmeli madde türünde yapılandırıldığı (Arslan ve Görgülü-Arı, 2025; Burkaz-Ekinci, 2021; Somuncu-Demir, 2012; Şeker, 2017) dikkate alındığında, geliştirilen başarı testi ile özel yetenekli öğrencilerin performanslarının çok boyutlu olarak değerlendirilmesi sağlanabilir. Bu noktada geliştirilen testin öne çıkan özelliği, öğrencilerin yalnızca temel kavrayış düzeylerini değil; açıklama yapma, gerekçelendirme, ilişkilendirme ve değerlendirme gibi üst düzey bilişsel süreçlerini de ortaya çıkarmalarını amaçlayan bir yapıya sahip olmasıdır. Özellikle açık uçlu maddelerin ilgili öğrenme çıktılarıyla eşleştirilmesi, sürdürülebilirlik temasının doğası gereği problem çözme ve değerlendirme süreçlerinin öğrenciler tarafından nasıl yapılandırıldığını ortaya koyabilir. Bu yönüyle geliştirilen test, özel yetenekli öğrenciler için öğretim hedeflerinde öne çıkan derinlik ve bilişsel karmaşıklık beklentisiyle uyumlu bir değerlendirme yaklaşımı sunarak alan yazındaki tek tür madde formatına dayalı ölçme uygulamalarına kıyasla daha kapsamlı bir değerlendirme olanağı sunabilir.

2024 Maarif Modeli Bilim ve Sanat Merkezleri Fen Bilgisi dersi çerçeve öğretim programında yer alan “Sürdürülebilirlik” ünitesi programda uzun bir zaman dilimine yayılan ve içerisinde farklı alt temalardan oluşan bir yapıya sahiptir. Alt düzey bilişsel beceriler, üst düzey düşünme becerileri için önkoşul niteliği taşıdığından; bu becerilerdeki yetersizlik, daha üst basamaklara geçişi zorlaştırabilir. Ya da tam tersi bir durumda öğrenciler üst düzeyde yer alan bilişsel becerilere sahiplerse alt düzeydeki becerilere ilgisiz kalabilirler. Bloom taksonomisinde yer alan bilişsel süreç becerileri birbirine bağlıdır ve bir basamakta tam olarak öğrenilemeyen bilgi diğer basamakları etkileyerek öğrenme sürecinde eksiklikler yaşanmasına neden olacaktır (Çakıcı vd., 2012). Bu nedenle, başarı testinde yer alan maddelerin, sürdürülebilirlik temasına ilişkin hem temel hem de üst düzey bilişsel becerileri kapsayacak şekilde ve öğrencilerin hazır bulunuşluk düzeylerindeki çeşitliliği yansıtacak şekilde yapılandırılması önemlidir.

Öneriler

Yapılan çalışma kapsamında aşağıdaki öneriler geliştirilmiştir:

- Özel yetenekli öğrencilerin üst düzey düşünme becerileri göz önüne alındığında sürdürülebilirlik temelli başarı testlerinde hem çoktan seçmeli hem de açık uçlu maddelerin birlikte kullanılması, öğrencilerin bu becerileri daha etkili şekilde ortaya koymalarına olanak sağlayabilir.
- Test kapsamının daraltılması veya madde sayısının azaltılması, öğrencilerin sürece daha istekli katılmasını sağlayabilir. Daha kısa ve odaklanmış bir test, özellikle yorulma etkisini azaltarak verilerin niteliğini artırabilir.
- Testin farklı sınıf düzeylerinde uygulanması, sürdürülebilirlikle ilgili kavramsal gelişimi yaş gruplarına göre izlemeye olanak tanıyabilir.
- Geliştirilen testin örneklem grubu Bilim ve Sanat Merkezi'nde öğrenim gören öğrencilerle sınırlı olduğundan, testin farklı okul ya da kurum türlerinde ve farklı öğrenci gruplarında uygulanarak bulguların karşılaştırılmasına yönelik araştırmalar yapılabilir.
- Bu çalışmada geçerlik kanıtları ağırlıklı olarak kapsam geçerliği ve madde analizi bulguları üzerinden sunulduğundan, ilerleyen çalışmalarda farklı örneklemlemlerle test puanlarının yapı geçerliğine ilişkin analizlerin yürütülmesi ve bu yolla geçerlik kanıtlarının çeşitlendirilmesi önerilebilir.

Kaynakça

- Akay, H., Soybaş, D., & Argün, Z. (2006). Problem kurma deneyimleri ve matematik öğretiminde açık-uçlu soruların kullanımı. *Kastamonu Eğitim Dergisi*, 14(1), 129-146.
- American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME]. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Arslan, K., & Görgülü Arı, A. (2025). Ortaokul öğrencilerine yönelik sürdürülebilirlik başarı testinin geliştirilmesi. *International Journal of Social Science Research*, 14(1), 126-150. <https://izlik.org/JA95BR64WM>
- Aslanoglu, A. E. (2022). Üst düzey zihinsel beceriler ve ölçülmesi. İ. Karakaya (Ed.), *Açık uçlu soruların hazırlanması, uygulanması ve değerlendirilmesi* içinde (1. Baskı, s.2-27). Pegem Akademi.
- Atalmış, E. H. (2022). Çoktan seçmeli testler. B. Çetin (Ed.), *Eğitimde ölçme ve değerlendirme* içinde (2. Baskı, s.188-216). Anı Yayıncılık.
- Ayre, C., & Scally, A. J. (2014). Critical values for Lawshe's content validity ratio: Revisiting the original methods of calculation. *Measurement and Evaluation in Counseling and Development*, 47(1), 79-86. <https://doi.org/10.1177/0748175613513808>
- Belen, B. (2020). Ortaöğretim öğrencilerinin sürdürülebilir çevre hakkındaki bilgi, tutum ve davranışlarının belirlenmesi (Tez No.612282) [Yüksek Lisans Tezi, Gazi Üniversitesi]. Yükseköğretim Kurulu Ulusal Tez Merkezi.
- Birgin, A., & Özcan, H. (2022). 8. sınıf öğrencilerinin mevsimlerin oluşumu ile ilgili bilgilerini ölçmeye yönelik bir başarı testinin geliştirilmesi. *Trakya Üniversitesi Sosyal Bilimler Dergisi*, 24(1), 305-326. <https://doi.org/10.26468/trakyasobed.817792>
- Brookhart, S. (2010). *How to assess higher-order thinking skills in your classroom*. ASCD.
- Burkaz-Ekinci, S. (2021). Ortaokul öğrencileri için sürdürülebilir kalkınma eğitimine yönelik bir modül geliştirme çalışması (Tez No. 699992) [Yayınlanmamış Doktora Tezi, Recep Tayyip Erdoğan Üniversitesi]. Yükseköğretim Kurulu Ulusal Tez Merkezi.
- Corash, D.N., & Jones, M. (2012). Keeping young gifted students engaged through science. *Understanding Our Gifted*, 25(1), 9-15.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Cengage Learning.
- Cross, T. L. (2011). *On the social and emotional lives of gifted children: Understanding and guiding their development* (4th ed.). Prufrock Press.
- Çakıcı, Y., Ürek, H., & Dinçer, E. O. (2012). İlköğretim öğrencilerinin soru oluşturma becerilerinin incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 8(1), 43-68.
- Çetin, B. (2010). Bilişsel alan davranışlarının ölçülmesi. M. Gömleksiz & S. Erkan (Ed.), *Eğitimde Ölçme ve Değerlendirme* içinde (2. Baskı, s. 71-143). Nobel Akademi Yayıncılık.

- Emir, S. (2021). *Üstün zekalı ve yetenekliler için program farklılaştırma*. Vize Akademik Yayıncılık.
- Gökmen, A., Solak, K., & Ekici, G. (2017). Sürdürülebilir kalkınma için eğitim: Öğretmen adaylarının tutumları ile ilişkili olan faktörler. *Kesit Akademi Dergisi*, 12, 462-480. <https://doi.org/10.18020/kesit.1341>
- Instructional Assessment Resources (IAR). (2011). *Interpreting test results*. <http://www.utexas.edu/academic/ctl/assessment/iar/students/report/itemanalysis.php>
- İlhan, M. (2022). Madde analizi ve madde ile test istatistikleri arasındaki ilişki. B. Çetin (Ed.), *Eğitimde ölçme ve değerlendirme* içinde (2. Baskı, s. 269-298). Anı Yayıncılık.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1-73. <https://doi.org/10.1111/jedm.12000>
- Karakaya, İ., & Şata, M. (2022). Açık uçlu maddeler. İ. Karakaya (Ed.), *Açık uçlu soruların hazırlanması, uygulanması ve değerlendirilmesi* içinde (1. Baskı, s. 28-39). Pegem Akademi.
- Karamustafaoğlu, S., & Tutar, M. (2016). 6. sınıf “Dünyamız, Ay ve yaşam kaynağımız Güneş” ünitesine yönelik bir başarı testi geliştirme. Ö. Demirel & S. Dinçer (Ed.), *Eğitim bilimlerinde yenilikler ve nitelik arayışı* içinde (s. 303-320). Pegem Akademi.
- Kızılkaya, A., & Seven, S. (2017). Fen öğretiminde Jigsaw I tekniğinin öğrencilerin Bloom taksonomisi’nin bilişsel alan alt ve üst düzey akademik başarılarına etkisi. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, 19(2), 250-270. <https://doi.org/10.17556/erziefd.334984>
- Kline, P. (1994). *An Easy guide to factor analysis*. Routledge.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Krathwohl, D. R. (2002). A Revision of Bloom’s taxonomy: An Overview. *Theory Into Practice*, 41(4), 212-218. https://doi.org/10.1207/s15430421tip4104_2
- Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 263-313). Emerald Group Publishing.
- Laverty, J. T., Underwood, S. M., Matz, R. L., Posey, L. A., Carmel, J. H., Caballero, M. D., Fata-Hartley, C. L., Ebert-May, D., Jardeleza, S. E., & Cooper, M. M. (2016). Characterizing college science assessments: The Three-dimensional learning assessment protocol. *PLoS ONE*, 11(9), 1-21. <https://doi.org/10.1371/journal.pone.0162333>
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28(4), 563-575. <https://doi.org/10.1111/j.1744-6570.1975.tb01393.x>
- Leu, D. J., Forzani, E., Burlingame, C., Kulikowich, J., Sedransk, N., Coiro, J., & Kennedy, C. (2013). The new literacies of online research and comprehension: Assessing and preparing students for the 21st century with Common Core State Standards. In S. B. Neuman & L. B. Gambrell (Eds.), *Quality reading instruction in the age of Common Core standards* (pp. 219-240). International Reading Association.
- McCollister, K., & Saylor, M. F. (2010). Lift the ceiling: Increase rigor with critical thinking skills. *Gifted Child Today*, 33(1), 41-47. <https://doi.org/10.1177/107621751003300110>
- Millî Eğitim Bakanlığı [MEB] (2024). *Türkiye Yüzyılı Maarif Modeli*. <https://tymm.meb.gov.tr/>
- Organisation for Economic Cooperation and Development (OECD). (2018). *The future of education and skills: Education 2030*. OECD Publishing.
- Popham, J. (2003) Trouble with testing: Why standard-based assessment doesn’t measure up. *American School Board Journal*, 190(2), 14-17.
- Silverman, L. K. (2003). *Characteristics of giftedness scale: Research and review of the literature*. Gifted Development Center. http://www.gifteddevelopment.com/PDF_files/scalersrch.pdf
- Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100-107. <https://doi.org/10.7334/psicothema2013.256>
- Somuncu-Demir, N. (2012). *Sürdürülebilir çevre eğitimi kapsamında gerçekleştirilen tarım uygulamalı bahçe temelli eğitim modelinin değerlendirilmesi* (Tez No. 317092) [Yayınlanmamış Doktora Tezi, Gazi Üniversitesi]. Yükseköğretim Kurulu Ulusal Tez Merkezi.
- Şahin, F. (2018). *Özel yeteneklilerin eğitimi: Eğitsel stratejiler ve örneklerle zenginleştirilmiş müfredat farklılaştırma modelleri*. Nobel Akademi Yayıncılık.
- Şata, M., & Karakaya, İ. (2022). Açık uçlu maddelerin analizi. İ. Karakaya (Ed.), *Açık uçlu soruların hazırlanması, uygulanması ve değerlendirilmesi* içinde (1.bsk., s.104-143). Pegem Akademi.

- Şeker, F. (2017). *Fen eğitiminde sürdürülebilirlik kavramının değerlendirilmesi ve model programın oluşturulması* (Tez No. 480009) [Yayınlanmamış Doktora Tezi, Kastamonu Üniversitesi]. Yükseköğretim Kurulu Ulusal Tez Merkezi.
- Tabachnick, B.G., & Fidell, L.S. (2001) *Using multivariate statistics* (4th Edition). Allyn and Bacon.
- Uddin, I., Uddin, I., Rehman, I. U., Siyar, M., & Mehboob, U. (2020). Item analysis of multiple choice questions in pharmacology. *Journal of Saidu Medical College Swat*, 10(2), 128–131. <https://doi.org/10.52206/jsmc.2020.10.2.320>
- Warne, R. T. (2012). History and development of above-level testing of the gifted. *Roeper Review*, 34(3), 183–193. <https://doi.org/10.1080/02783193.2012.686425>
- Warne, R. T. (2014). Using above-level testing to track growth in academic achievement in gifted students. *Gifted Child Quarterly*, 58(1), 3–23. <https://doi.org/10.1177/0016986213513793>

EXTENDED SUMMARY

The rapid emergence of new technologies, digital transformation, and intensified global interaction in the 21st century has significantly reshaped how individuals learn. This shift, driven by accelerated access to information, has exposed the limitations of traditional educational structures while fostering a transition toward more comprehensive, process-oriented, and evidence-based assessment approaches (Brookhart, 2010; Lavery et al., 2016). Contemporary assessment frameworks aim not only to determine what students know but also to examine how they process, interpret, and transfer knowledge across contexts. This evolution necessitates the development of measurement tools that incorporate diverse item formats capable of capturing both foundational understanding and higher-order cognitive processes. Sustainability, as an inherently interdisciplinary and multifaceted construct, requires such dual-format assessment designs. While multiple-choice items can efficiently measure conceptual understanding, open-ended items can capture reasoning, interpretation, and creative problem-solving. These skills are especially salient among gifted students, who tend to integrate complex information and generate original solutions at advanced levels (Cross, 2011; Silverman, 2003). The integration of the Sustainability unit into the Science curriculum of Science and Art Centers under the Türkiye Yüzyılı Maarif Model (MEB, 2024) has further underscored the need for a valid and reliable assessment tool capable of evaluating gifted students' competencies within this expanding domain. Addressing this need, the present study developed an academic achievement test for fourth- and fifth-grade gifted students by following the systematic test development procedures outlined by Crocker and Algina (2006). Learning outcomes were aligned with the national curriculum, content validity was supported through a table of specifications and expert review, and item formats were structured in accordance with the Revised Bloom's Taxonomy to represent a range of cognitive process levels. Expert evaluations were analyzed using the Lawshe technique, and items that failed to meet the minimum content validity ratio were either eliminated or revised. Following ethical approval and the required institutional permissions, a pilot implementation was conducted to assess the clarity, comprehensibility, and feasibility of the test form. Based on the feedback obtained from the pilot implementation, the revised preliminary form was administered to 238 gifted students from various regions of Türkiye. Multiple-choice items were analyzed using the Test Analysis Program (TAP), whereas open-ended responses were independently scored by two raters. Inter-rater reliability was assessed using the Intraclass Correlation Coefficient (ICC). The finalized test consisted of 29 multiple-choice items and nine open-ended questions. Because the first open-ended question included four separately scored subquestions, the open-ended section comprised 12 scorable response units. The maximum possible score on the final test was 53.

The multiple-choice section demonstrated an average item difficulty index of 0.55 and an average discrimination index of 0.47. Reliability analyses yielded KR-20 and KR-21 coefficients of 0.81 and 0.80, respectively, indicating strong internal consistency. For the open-ended items, item difficulty indices ranged from 0.22 to 0.62, whereas discrimination indices ranged from 0.35 to 0.72. Inter-rater reliability was $ICC = 0.979$ at the item level and $ICC = 0.989$ for total scores, indicating exceptionally high agreement between raters. The item difficulty and discrimination findings further indicated that the test was able to assess both foundational conceptual understanding and higher-order cognitive processes.

Collectively, these psychometric findings demonstrate that the developed assessment tool can produce valid and reliable measurements of gifted students' knowledge and skills related to sustainability. Beyond its psychometric strengths, the test offers practical value for instructional monitoring, the evaluation of differentiated teaching practices, and curriculum development efforts. By incorporating multiple-choice and open-ended items aligned with Bloom's taxonomy, particularly those addressing higher-order thinking skills, the test provides an opportunity to evaluate the depth and complexity of understanding characteristic of gifted learners. The findings suggest that this tool may serve as a meaningful feedback mechanism for identifying students' strengths and learning needs, guiding instructional refinement, and informing enrichment practices. Future research may further examine the test's psychometric properties using larger and more diverse samples and explore the adaptation of similar assessment frameworks to other interdisciplinary units within gifted education programs.

YAZARLARIN BEYANI / DECLARATION OF THE AUTHORS

Katkı Oranı Beyanı	☐ Yazar, çalışmanın tümüne tek başına katkı sağlamıştır. ☒ Yazarlar çalışmaya eşit oranda katkı sağlamıştır. ☐ Çalışmaya birinci yazar %XX oranında, ikinci yazar %XX oranında, üçüncü yazar %XX oranında katkı sağlamıştır.
Declaration of Contribution Rate	☐ The author contributes the study on his/her own ☒ The authors have equal contributions. ☐ The first author contributes XX%, the second author contributes XX%, the third author contributes XX%.
Çatışma Beyanı	☒ Çalışmada herhangi bir potansiyel çıkar çatışması söz konusu değildir. ☐ Yazarlar, çalışmadan XXX (firması, kurumu, kişinin) etkilenebileceğini ve ortaya çıkacak çıkar çatışması durumunu yönetmek için onaylanmış bir planının olduğunu beyan etmektedir.
Declaration of Conflict	☒ There is no potential conflict of interest in the study. ☐ Authors declare that XXX (firm, institution, person) may be affected from the study and in case of a conflict of interest they have a confirmed plan in order to administer the case.
Yayın Etiği Beyanı	☒ Çalışmada etik dışı bir husus bulunmadığını, araştırma ve yayın etiğine özenle uyulduğunu beyan ederiz. ☐ Bu çalışma, etik kurul belgesi gerektiren bir çalışma değildir. ☒ Bu çalışma için Milli Eğitim Bakanlığı'ndan 29.05.2025 tarihli ve MEB.TT.2025.025626.01 sayılı karar ile etik kurul onayı alınmıştır.
Declaration of Publication Ethics	☒ We hereby declare that the study has not unethical issues and that research and publication ethics have been observed carefully. ☐ This study does not require ethics committee approval. ☒ Ethical committee approval for this study was obtained from the Ministry of National Education with decision number MEB.TT.2025.025626.01 dated 29.05.2025.
Üretken Yapay Zekâ Kullanım Beyanı	☒ Bu çalışmanın hiçbir aşamasında üretken yapay zekâ araçlarından faydalanılmamıştır. ☐ Bu çalışmanın hazırlanması sırasında yazar(lar), [NEDEN] amacıyla [ARAÇ/HİZMET ADI] kullanmıştır. Bu araç/hizmeti kullandıktan sonra, yazar(lar) içeriği gerektiği şekilde gözden geçirip düzenlemiş ve yayınlanan makalenin içeriği konusunda tüm sorumluluğu üstlenmiştir.
Declaration of Generative AI Use	☒ No generative artificial intelligence tools were used at any stage of this study. ☐ During the preparation of this study the author(s) used [NAME OF TOOL / SERVICE] in order to [REASON]. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.
Destek ve Teşekkür Beyanı	☒ Çalışmada herhangi bir kurum ya da kuruluştan mali destek alınmamıştır. ☐ Bu çalışma, X Projesi (Proje No: XXXXX) tarafından desteklenmektedir. ☐ Diğer (Belirtiniz)
Declaration of Support and Acknowledgments	☒ No financial support is taken from any institution or organization. ☐ The study is supported by X Project (Project No: XXXXX). ☐ Others (Specify)