

'Aynalarda Yansıma ve Işığın Soğrulması' Konusunda Geçerliliği ve Güvenilirliği Sağlanmış Başarı Testi Geliştirme Çalışması

A Study on Developing a Reliable and Valid Achievement Test on 'Light Reflection in Mirrors and Light Absorption' Subject*

Aslı SAYLAN-KIRMIZIGÜL** 

Hasan KAYA*** 

Received: 24 May 2018

Research Article

Accepted: 04 February 2019

ABSTRACT: The purpose of this study is to develop an achievement test on "Light Reflection in Mirrors and Light Absorption" unit in seventh grade curriculum by taking into account the test development steps. For this purpose, 30 multiple-choice and 7 open-ended questions were prepared in accordance with the learning outcomes. The achievement test was applied 180 seventh grade students from two middle schools in Kayseri during the spring semester of 2016-2017 academic years. The final form of the achievement test has totally 36, 29 multiple-choice and 7 open-ended, questions. The reliabilities of multiple-choice and open-ended questions were 0.82 and 0.78, respectively; whereas the mean of Kappa coefficients of open-ended questions was 0.82. The results show that it is valid and reliable to use the achievement test in scientific studies or classroom assessment processes, and it also can be used as pre-test and post-test by researchers who carry out scientific research in the field of science education.

Keywords: achievement test, test development, validity, reliability, science education.

ÖZ: Bu çalışmanın amacı test geliştirme basamakları dikkate alınarak 7. sınıf öğrencilerinin, öğretim programında bulunan "Aynalarda Yansıma ve Işığın Soğrulması" ünitesine yönelik başarılarını ortaya koyan geçerli ve güvenilir bir başarı testi geliştirmektir. Bu amaçla, mevcut ünite kazanımlarına uygun olarak 30 çoktan seçmeli ve 7 açık uçlu soru maddesi hazırlanmıştır. Başarı testi 2016-2017 öğretim yılının bahar döneminde, Kayseri'de bulunan iki ortaokulda öğrenim görmekte olan 180 yedinci sınıf öğrencisine uygulanmıştır. Nihai başarı testi 29 çoktan seçmeli soru ve 7 açık uçlu soru olmak üzere toplamda 36 soru içermektedir. Nihai testte bulunan çoktan seçmeli soruların güvenilirliği 0.82, açık uçlu soruların güvenilirliği 0.78 olarak, açık uçlu soruların Kappa katsayılarının ortalama değeri ise 0.82 olarak bulunmuştur. Elde edilen bulgular başarı testinin bilimsel çalışmalarda ve sınıf içi değerlendirme süreçlerinde kullanılması geçerli ve güvenilir bir ölçme aracı olduğunu göstermektedir. Ayrıca geliştirilmiş olan bu test fen eğitimi alanında bilimsel çalışmalar gerçekleştiren araştırmacılar tarafından ön test ve son test olarak da kullanılabilir.

Anahtar kelimeler: başarı testi, test geliştirme, geçerlilik, güvenilirlik, fen eğitimi.

* Bu çalışma, birinci yazarın Erciyes Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından SDK-2017-7115 kodlu proje ile desteklenen "Fen Eğitiminde Bilgisayar Destekli, Etkinlik Temelli ve Sorgulamaya Dayalı Öğretim Yaklaşımlarının Karşılaştırılması" başlıklı doktora tez çalışmasının bir bölümünü içermektedir.

** *Corresponding Author:* Dr., Erciyes University, Kayseri, Turkey, aslisaylan@erciyes.edu.tr, <https://orcid.org/0000-0001-5678-8050>

*** Prof. Dr., Erciyes University, Kayseri, Turkey, hasankaya@erciyes.edu.tr, <https://orcid.org/0000-0003-3529-9762>

Citation Information

Saylan-Kırmızıgül, A., & Kaya, H. (2019). 'Aynalarda yansıma ve ışığın soğrulması' konusunda geçerliliği ve güvenilirliği sağlanmış başarı testi geliştirme çalışması. *Kuramsal Eğitim Bilim Dergisi [Journal of Theoretical Educational Science]*, 12(2), 474-493.

Giriş

İçinde bulunduğumuz bilim ve teknoloji çağında, fen bilimleri öğretiminin önemi gün geçtikçe artmaktadır. Fen bilimleri dersi hayatı anlamlandırmada önemli bir yer tutmaktadır. Türkiye yıllardır PISA ve TIMSS gibi uluslararası sınavlarda özellikle fen bilimlerinde istenilen düzeye gelememektedir (Milli Eğitim Bakanlığı [MEB], 2013, 2014, 2016a, 2016b). Bu durum, ülkemizde fen eğitiminin istenilen düzeyde gerçekleştirilemediğini göstermektedir. Fizik, kimya ve biyoloji gibi farklı alt bilim dallarına ayrılan fen bilimleri dersinin içeriğinde soyut konu ve kavramların yer alması öğrencilerin bu derse karşı olumsuz tutum oluşturmalarına ve motivasyonlarının düşmesine neden olabilmektedir (Kutu, 2011; Uzun, 2013). Bu durum dersteki akademik başarının yetersiz olmasına yol açabilmektedir. Fizik, fen bilimleri içinde çok önemli bir yere sahiptir. Pek çok araştırma, fen bilimlerinin fizik konularındaki başarının kimya ve biyoloji konularına oranla daha düşük olduğunu, fiziğin olumsuz tutum geliştirilen, anlamakta güçlük çekilen ve başarısız olunan derslerin başında geldiğini göstermektedir (Bahar & Polat, 2007; Bakaç & Kumru 1998; Karamustafaoğlu, Bacanak, Değirmenci, & Karamustafaoğlu, 2010). Ortaöğretim ve yükseköğretim öğrencilerinin özellikle ışık ve görüntü oluşumu gibi günlük olayları açıklamakta zorluk çektikleri görülmektedir (Colin & Viennot, 2001; Galili & Hazan, 2000; Kocakülâh & Şardağ, 2013). Ortaokul öğrencilerinin "Işık ve Ses" ünitesi ile ilgili kavram yanlışlarının olduğu ve öğrencilerinin büyük bir kısmının ışığın tanımlanması, ışığın yayılması, ışığın yansıması ve kırılması kavramlarını anlamakta ve ifade etmekte güçlük çektiği görülmektedir (Akdeniz, Yıldız, & Yiğit, 2001; Şen, 2001). Fen bilimleri öğretim programı sarmal bir yapıya sahiptir. Bu durum, öğrencilerin aynı konu ile ilgili kavramları her yıl artan bir yoğunlukla öğrenmelerini amaçlamakta ve öğrencilerin ilk yıllarda sahip oldukları kavram yanlışlarını sonraki yıllara da taşımalarına, hatta diğer fizik kavramlarında da bazı yanlışların oluşmasına neden olmaktadır (Özsevgeç, Çepni, & Bayri, 2007).

Öğrencilerin öğrenme eksikliklerinin giderilmesinde ve derslerde verilen bilgilerin öğrenciler tarafından ne ölçüde kazanıldığına tespit edilebilmesinde ölçme ve değerlendirmenin önemli bir rolü bulunmaktadır. Sağlıklı bir ölçme ve değerlendirmenin yapılabilmesi için ise geçerlilik ve güvenilirliği sağlanmış ölçme araçlarına ihtiyaç duyulmaktadır (Gönen, Kocakaya, & Kocakaya, 2011). Bilgi aktarımına dayalı geleneksel öğretimden, öğretmenin rehber olduğu ve öğrencilerin süreçte aktif rol aldıkları yapılandırmacı yaklaşıma geçilmesi beraberinde yeni ölçme araçlarının geliştirilmesini de zorunlu kılmıştır (Brooks, 2002; Cochran Smith, 2002). Yapılandırmacı yaklaşımda öğrencilerin yalnızca çoktan seçmeli sorular ile değerlendirilmesinden ziyade açık uçlu sorular, gelişim dosyaları, performans ve proje görevleri gibi ölçme yöntemleri ile öğrenci başarısının değerlendirilmesi önerilmektedir (Uzunöz & Buldan, 2012). Buradan hareketle, bu çalışmada çoktan seçmeli ve açık uçlu sorulardan oluşan bir başarı testi geliştirilmesi amaçlanmıştır.

Ortaokul öğrencilerinin ışık konusuna ilişkin akademik başarılarının ölçülmesi amacı ile daha önce birtakım başarı testleri geliştirilmiştir (Güneş-Koç, 2013; Küçük, 2014; Sayın, 2015). Ancak 2015 yılında yenilenen müfredat ile birlikte ünitenin adı, kazanımları ve içeriği değiştiğinden, mevcut başarı testlerinin uygulanması konusunda birtakım sınırlılıklar ortaya çıkmıştır. Bu bağlamda, mevcut çalışma kapsamında yeni

bir başarı testinin geliştirilmesine karar verilerek kavram yanılgılarının sıklıkla gözlemlendiği ve öğrencilerin öğrenmekte güçlük çektikleri yeni öğretim programındaki adı ile “Aynalarda Yansıma ve Işığın Soğurulması” ünitesine yönelik bir başarı testi tasarlanmıştır. Dolayısı ile bu çalışmanın amacı test geliştirme basamakları dikkate alınarak 7. sınıf öğrencilerinin “Aynalarda Yansıma ve Işığın Soğurulması” konusundaki başarılarını ortaya koyan geçerli ve güvenilir bir ölçme aracı geliştirmektir.

Yöntem

Araştırma Deseni

Çalışmada nicel araştırma yöntemlerinden tarama desenine başvurulmuştur. Bir grubun belirli özelliklerinin belirlenmesi için verilerin toplanmasını amaçlayan çalışmalara tarama araştırması denir (Fraenkel, Wallen, & Hyun, 2012). Tarama desenindeki çalışmalarda örneklemden toplanan veriler evrenin eğilimleri, fikirleri veya tutumları hakkında nicel ve sayısal betimlemeler yapılmasını sağlar (Fraenkel, Wallen, & Hyun, 2012). Bu bağlamda, mevcut araştırmada belirli basamaklar takip edilerek bir başarı testi geliştirilmiştir. Öncelikle testin amacı ve ölçülecek kazanımlar belirlenmiştir. Daha sonra maddeler yazılarak gözden geçirilmiş ve deneme testi hazırlanmıştır. Pilot uygulama gerçekleştirilerek sonuçlar puanlandırılmıştır. İstatistiksel hesaplamalar sonucu teste son hali verilmiştir.

Örneklem

Araştırmanın ulaşılabilir evrenini 2016-2017 öğretim yılında Kayseri ilinde bulunan ortaokullarda öğrenim görmekte olan 7. sınıf öğrencileri oluşturmaktadır. Bu çalışmada örneklem kolay ulaşılabilir örnekleme yolu ile seçilmiştir. Başarı testi 2016-2017 öğretim yılı bahar döneminde, Kayseri'nin Talas merkez ilçesindeki iki ortaokulun 7. sınıf şubelerinde öğrenim görmekte olan, ünite kavram ve konularını öğrenmiş 180 öğrenciye uygulanmıştır. Madde analizinin yapılabilmesi için ölçme aracındaki madde sayısının en az 5 katı kadar örnekleme ihtiyaç vardır (Nunnally, 1967). Bu çalışmada 30 maddeden oluşan çoktan seçmeli soruların madde analizi için en az 150 kişi olması gerektiğinden, pilot çalışmada örneklem grubu 180 kişiden oluşmaktadır.

Veri Toplama Aracı

7. sınıf öğrencilerinin “Aynalarda Yansıma ve Işığın Soğurulması” konusundaki başarılarını ortaya koyacak geçerli ve güvenilir bir ölçme aracı geliştirmeyi amaçlayan bu çalışmada veri toplama aracı olarak çoktan seçmeli ve açık uçlu sorulardan oluşan bir başarı testi kullanılmıştır. Başarı testi geliştirilirken Marzano ve Kendall (2007) tarafından oluşturulmuş olan taksonomiden faydalanılmıştır. Marzano ve Kendall'ın (2007) bu sınıflandırmasının temelinde, hedeflerin hem zorlayıcı, hem de anlaşılır olması için farklı zorluk seviyelerinde bulunmaları gerektiği düşüncesi vardır. Taksonomiye göre üç temel sistem vardır: 1. Bilişsel sistem, 2. Biliş ötesi system, 3. Benlik sistemi. Bu çalışmada temel alınan bilişsel sistemde kolaydan zora doğru dört temel seviye bulunmaktadır: Geri getirme, kavrama, analiz, bilgiden yararlanma. Dört seviyenin her biri kendi içerisinde daha küçük bilişsel süreç kategorilerine ayrılmıştır. Veri toplama aracı için oluşturulan belirtke tablosu Tablo 1’de verilmiştir.

Tablo 1

Marzano ve Kendall (2007) Taksonomisi Kazanım Belirtke Tablosu

Bilişsel Sistem Seviyeleri		Testler ve Maddeleri	
		Çoktan Seçmeli	Açık Uçlu
Geri Getirme	Farkına varma	5, 12, 14	
	Hatırlama	6, 19	1a, 1b, 1c, 1d, 1e, 1f
	Gerçekleştirme	13	
Kavrama	Bütünleştirme	7, 8, 23	
	Sembolleştirme	15	2, 3, 7d
	Eşleştirme	4, 27	
Analiz	Sınıflandırma	9	6a, 6b, 6c
	Hataları analiz etme	10, 11, 20, 24, 25	
	Genelleme	2, 17, 26, 28, 29	4
	Belirtme		
Bilgiden Yararlanma	Karar verme	1, 16, 21, 30	
	Problem çözme	3, 18, 22	
	Deney yapma		7a, 7b, 7c
	Araştırma		5

Kapsam geçerliliğinin sağlanması için soruların taksonomi basamaklarına ek olarak fen bilimleri öğretim programında belirlenmiş olan kazanımlara da uygun olarak hazırlanmasına özen gösterilmiştir. Ünite kazanımları şunlardır:

1. Ayna çeşitlerini gözlemler ve kullanım alanlarına örnekler verir.
2. Düz, çukur ve tümsek aynalarda oluşan görüntüleri karşılaştırır. [Çukur aynada cismin görüntüsünün özelliklerinin (büyük/küçük, ters/düz) cismin aynaya olan uzaklığına göre değişebileceği belirtilir.]
3. Işığın madde ile etkileşimi sonucunda madde tarafından soğrulanabileceğini keşfeder.
4. Beyaz ışığın tüm ışık renklerinin bileşiminden oluştuğu sonucunu çıkarır.
5. Gözlemleri sonucunda cisimlerin, siyah, beyaz ve renkli görünmesinin nedenini, ışığın yansıması ve soğrulmasıyla ilişkilendirir.
6. Güneş enerjisinin günlük yaşam ve teknolojideki yenilikçi uygulamalarına örnekler verir ve kaynakların etkili kullanımı bakımından Güneş enerjisinin önemini tartışır.

2018 yılı fen bilimleri öğretim programında, yukarıda bahsi geçen ünitenin adı "Işığın Madde ile Etkileşimi" olarak değiştirilmiş ve yukarıdaki kazanımlar aynen korunmuştur (MEB, 2018). Bu kazanımlara ek olarak ışığın kırılması ve mercekler konusu ile ilgili aşağıda verilen beş kazanım üniteye dâhil edilmiştir:

1. Ortam değiştiren ışığın izlediği yolu gözlemleyerek kırılma olayının sebebini ortam değişikliği ile ilişkilendirir.

2. Işığın kırılmasını, ince ve kalın kenarlı mercekler kullanarak deneyle gözlemler.

3. İnce ve kalın kenarlı merceklerin odak noktalarını deneyerek belirler.

4. Merceklerin günlük yaşam ve teknolojiadaki kullanım alanlarına örnekler verir.

5. Ayna veya mercekleri kullanarak bir görüntüleme aracı tasarlar.

Güncellenen kazanımlara uygun olarak, başarı testinin çoktan seçmeli sorulardan oluşan bölümü için literatürde yer alan başarı testlerinden (Güneş-Koç, 2013; Küçük, 2014; Sayın, 2015) 7 soru, MEB 7. sınıf fen bilimleri kitabından 6 soru (MEB, 2015); MEB Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü tarafından hazırlanmış olan 7. sınıf fen bilimleri kazanım kavrama testlerinden 3 soru, 2009 ve 2011 SBS 7. sınıf ve 2007 OKS 8. sınıf fen bilimleri sorularından 1'er soru teste dâhil edilmiştir (MEB, 2007, 2009, 2011). Sonuç olarak bahsedilen kaynaklardan yararlanılarak oluşturulan 19 soru ve araştırmacılar tarafından yazılan 11 soru birleştirilerek "Aynalarda Yansıma ve Işığın Soğrulması" ünitesi kazanımlarına uygun dört seçenekli 30 soruluk çoktan seçmeli bir başarı testi hazırlanmıştır.

Açık uçlu sorular, katılımcıların serbestçe cevap vermeleri istendiğinde kullanılmaktadır. Araştırmacı bu sorular sayesinde beklenen cevapların yanı sıra beklemediği cevapları da alarak konu hakkında daha geniş ve ayrıntılı bilgiler elde edebilir. Açık uçlu sorular cevaplandırılma şekline göre yorumlama, listeleme ve boşluk doldurma olmak üzere üç grupta toplanmaktadır (Mertens, 2005). Bu çalışmada öncelikle 14 açık uçlu sorudan oluşan bir soru havuzu hazırlanmıştır. Daha sonra bu soruların içinden seçilen 7 sorunun uygulanmasına karar verilmiştir. Bu sorulardan 5'i araştırmacı tarafından, 2'si ise farklı kaynaklardan yararlanılarak hazırlanmıştır (Güneş-Koç, 2013; TIMSS, 2011). Soruların açık, anlaşılır, kazanımlara uygun, farklı tiplerde (yorumlama, listeleme, boşluk doldurma) olmasına dikkat edilmiştir.

Hazırlanan çoktan seçmeli ve açık uçlu soruların 2016 yılında güncellenen öğretim programı kazanımlarına göre dağılımı Tablo 2'de verilmiştir.

Tablo 2

Soru Maddelerinin Kazanımlara Göre Dağılımı

Kazanımlar	Testler ve Maddeleri	
	Çoktan Seçmeli	Açık Uçlu
1	1, 6, 10, 18, 25	1a, 1b, 1c, 2, 6a, 6b
2	3, 9, 15, 22, 27	3, 6c
3	2, 7, 13, 21, 29	7
4	5, 12, 19, 24, 30	4
5	8, 14, 17, 23, 28	1d, 1e, 1f
6	4, 11, 16, 20, 26	5

Sonuç olarak, 7. sınıf öğrencilerinin Aynalarda Yansıma ve Işığın Soğrulması ünitesine yönelik akademik başarılarının ölçülmesi amacı ile 30 çoktan seçmeli ve 7 açık uçlu sorudan oluşan 37 soruluk bir akademik başarı testi geliştirilmiştir. Başarı testi

iki bölümden oluşmaktadır. Birinci bölümde bulunan 30 adet çoktan seçmeli soru maddesinden seçilen örnekler Ek-1'de, ikinci bölümdeki 7 adet açık uçlu soru maddesinden seçilen örnekler Ek-2'de verilmiştir. Hazırlanan başarı testi, kapsam geçerliliğinin sağlanması için deneyimli bir fen bilimleri öğretmeni ve iki akademisyenden oluşan 3 kişilik uzman grubunun görüşlerine sunulmuştur. Uzmanların görüş ve önerileri doğrultusunda çeşitli düzeltmeler yapılarak test pilot çalışma için uygulanabilir hale getirilmiştir ve 180 7. sınıf öğrencisine uygulanmıştır. Pilot testin ardından teste son hali verilmiştir.

Verilerin Analizi

Araştırmada, başarı testinin ölçtüğü özelliği ne derece duyarlı ölçebildiği, sonuçların hatalardan ne derece arındığının belirlenmesi için güvenilirlik analizi yapılmıştır. Ayrıca testin ölçülmesi amaçlanan özelliği başka bir özellikle karıştırmadan doğru bir şekilde ölçebilme derecesinin belirlenmesi için geçerlilik analizi yapılmıştır. Testte bulunan çoktan seçmeli ve açık uçlu sorular için farklı analizler gerçekleştirilmiştir.

Çoktan seçmeli başarı testinden elde edilen nicel verilerin analizinde MS Excel 2013, ITEMAN 4.2.1, SPSS 24 ve FACTOR 10.8.04 programları kullanılmıştır. Testin güvenilirlik ve geçerlilik çalışmaları kapsamında betimsel istatistikler, madde analizi ve tetrakorik korelasyon matrisi üzerinden açılımlı faktör analizi (AFA) teknikleri kullanılmıştır. Bu kapsamda, öncelikle MS Excel 2013 ile ITEMAN 4.2.1. programları birbirini teyit amaçlı olarak birlikte kullanılmış ve ABT-1 için madde analizi gerçekleştirilerek madde güçlük ve ayırt edicilik indeksleri hesaplanmıştır. Kavram yanılgısı testleri ve performans ölçen testler ölçüte dayalı (criterion-referenced), başarı düzeyini ölçen testler ise sıralamaya dayalı (norm-referenced) testler olduğundan (Gronlund & Linn, 1990), madde analizinde hesaplanan madde güçlük ve ayırt edicilik indeksleri sıralamaya dayalı testlere uygun olarak yorumlanmıştır. Madde analizinin ardından KR-20 yöntemi ile güvenilirlik analizi yapılmış ve ardından SPSS programı kullanılarak betimsel istatistikler ve Cronbach Alpha güvenilirlik katsayısı hesaplanmıştır. Son olarak, FACTOR programı kullanılarak tetrakorik korelasyon matrisi üzerinden AFA gerçekleştirilerek çoktan seçmeli teste ait faktörler belirlenmiştir.

Açık uçlu soruların analizinde de yine SPSS 24 programı kullanılarak Cronbach Alpha güvenilirlik katsayısı hesaplanmıştır. Daha sonra, 180 öğrencinin arasından rastgele seçilen 10 öğrencinin cevap kâğıtları araştırmacı tarafından geliştirilmiş olan rubrik (dereceli puanlama anahtarı) kullanılarak fen eğitiminde öğretim üyesi olan iki uzman tarafından bağımsız olarak puanlandırılmış ve gözlemcilerin arasındaki tutarlılığa bakılarak Kappa (κ) katsayısı hesaplanmıştır. Bu katsayı basit bir şekilde meydana gelmeyen şans beklentili uzlaşmazlıkların bir oranıdır (Cohen, 1960). Puanlayıcıların yaptıkları bu puanlamalar arasındaki farkların karşılaştırılması amacıyla, hem her bir madde için ayrı ayrı hem de tüm maddelerin toplam puanlarının ortalamaları arasındaki farklara ilişkin t-testi yapılmıştır.

Bulgular

Çoktan Seçmeli Sorulara İlişkin Pilot Uygulama ve Analizler

Madde analizi. Başarı testinde bulunan çoktan seçmeli her bir sorunun yalnız bir doğru cevabı bulunmaktadır. Analizlerde her bir doğru cevap için 1, yanlış cevap için ise 0 puan verilerek puanlama yapılmıştır. Yanlış cevaplar doğru cevapları etkilememektedir. Madde analizi için öncelikle, öğrencilerin başarı testine verdikleri cevaplar puanlanmıştır. Böylece her öğrencinin bir test skoru olmuştur. Daha sonra en düşük puanı alan öğrenciden başlanarak en yüksek puanı alan öğrenciye doğru bir sıralama yapılmıştır. Katılımcıların %27'si hesaplanarak yüksek puan alan öğrencilerin 49'unun üst grupta, düşük puan alan öğrencilerin 49'unun alt grupta yer aldığı belirlenmiştir.

Madde güçlük indeksinin (P) 0'a yakın olması maddenin zor olduğunu, 1'e yakın olması ise maddenin kolay olduğunu göstermektedir (Crocker & Algina, 1986). Her bir madde için P değerinin 0.40-0.60 aralığında olması tercih edilmektedir. Bir testin ortalama madde güçlük indeksinin de bu aralıkta olması gerekmektedir (Hotiu, 2006). Genel olarak, P değeri 0.20-0.90 arasında olan maddeler iyi ve kabul edilebilir olarak değerlendirilmektedir. P değeri bu aralığın dışında kalan maddeler testten çıkarılmalıdır (Boopathiraj & Chellamani, 2013). Bu çalışmada, tüm maddelerin güçlük indeksleri 0.20-0.90 aralığında bulunmuştur (Tablo 3). Yalnızca güçlük indeksleri temel alındığında testten çıkarılması gereken herhangi bir madde bulunamamıştır. Maddelerin ortalama güçlük indeksi 0.59 olduğundan, test güvenilirirdir.

Tablo 3

Madde Güçlük (P) ve Ayırt Edicilik (r) İndeksleri

Madde	D_u	D_a	P	Güçlük	r	Ayırt edicilik	Sonuç
1	37	23	0.61	kolay	0.29	zayıf	düzeltilmeli
2	43	20	0.64	kolay	0.47	çok iyi	kalmalı
3	17	27	0.45	orta	-0.20	çok zayıf	çıkarılmalı
4	38	7	0.46	orta	0.63	çok iyi	kalmalı
5	47	25	0.73	kolay	0.45	çok iyi	kalmalı
6	44	20	0.65	kolay	0.49	çok iyi	kalmalı
7	44	17	0.62	kolay	0.55	çok iyi	kalmalı
8	49	11	0.61	kolay	0.78	çok iyi	kalmalı
9	42	4	0.47	orta	0.78	çok iyi	kalmalı
10	29	10	0.40	orta	0.39	iyi	kalmalı
11	44	19	0.64	kolay	0.51	çok iyi	kalmalı
12	47	22	0.70	kolay	0.51	çok iyi	kalmalı
13	44	29	0.74	kolay	0.31	iyi	kalmalı

D_u : Maddeyi üst grupta doğru cevaplayan öğrencilerin sayısı

D_a : Maddeyi alt grupta doğru cevaplayan öğrencilerin sayısı

Tablo 3

Devamı

Madde	$D_{\bar{u}}$	D_a	P	Güçlük	r	Ayırt edicilik	Sonuç
14	42	21	0.64	kolay	0.43	çok iyi	kalmalı
15	30	20	0.51	orta	0.20	zayıf	düzeltilmeli
16	44	24	0.69	kolay	0.41	çok iyi	kalmalı
17	41	17	0.59	orta	0.49	çok iyi	kalmalı
18	43	20	0.64	kolay	0.47	çok iyi	kalmalı
19	28	18	0.47	orta	0.20	zayıf	düzeltilmeli
20	38	20	0.59	orta	0.37	iyi	kalmalı
21	46	16	0.63	kolay	0.61	çok iyi	kalmalı
22	44	13	0.58	orta	0.63	çok iyi	kalmalı
23	36	13	0.50	orta	0.50	çok iyi	kalmalı
24	37	13	0.51	orta	0.49	çok iyi	kalmalı
25	36	22	0.59	orta	0.29	zayıf	düzeltilmeli
26	44	24	0.69	kolay	0.41	çok iyi	kalmalı
27	39	16	0.56	orta	0.47	çok iyi	kalmalı
28	39	16	0.56	orta	0.47	çok iyi	kalmalı
29	45	18	0.64	kolay	0.55	çok iyi	kalmalı
30	41	21	0.63	kolay	0.29	zayıf	düzeltilmeli

$D_{\bar{u}}$: Maddeyi üst grupta doğru cevaplayan öğrencilerin sayısı

D_a : Maddeyi alt grupta doğru cevaplayan öğrencilerin sayısı

Bir testte soruların başarılı öğrenciler (üst grup) tarafından daha yüksek, başarısız öğrenciler (alt grup) tarafından ise daha düşük oranda doğru cevaplandırılması beklenir. Madde ayırt edicilik indeksinin (r) 0'a yaklaşması o sorunun üst ve alt grubu ayırt ediciliğinin düşük, +1'e yaklaşması ayırt ediciliğinin yüksek olduğunu gösterir. r 'nin negatif değerler alması sorunun doğru cevaplanma oranının alt grupta daha yüksek olması anlamına gelir. Böyle bir madde testin amacına hizmet etmemekte, aksine test güvenilirliğini düşürmektedir (Kubiszyn & Borich, 2003). Dolayısı ile r değeri sıfır veya negatif olan maddeler teste dahil edilmemelidir. r değeri 0.40 veya daha yüksek ise madde çok iyidir, aynen kullanılabilir; 0.30-0.39 arasında ise madde iyidir, düzeltilmesi gerekmez; 0.20-0.29 arasında ise madde zorunlu hallerde aynen kullanılabilir ancak bir miktar değiştirilmesi tercih edilir; 0.20'den küçük ise soru testten çıkarılmalıdır (Ebel & Frisbie, 1986). Pilot uygulamaya tabi tutulan başarı testinin 3. sorusuna ait ayırt edicilik indeksi -0.20'dir (Tablo 3). Bu değer negatif olması maddenin başarılı ve başarısız öğrencileri birbirinden ayırt etmediğini göstermektedir. Nitekim bu soru maddesini alt grupta doğru cevaplayan öğrenci sayısı üst grupta doğru cevaplayan öğrenci sayısından fazladır. Dolayısı ile sorunun testten çıkarılmasına karar verilmiştir. Diğer taraftan, 15. ve 19. maddelerin ayırt edicilik indeksleri 0.20; 1. ve 25. maddelerin ayırt edicilik indeksleri ise 0.29 olduğundan, bu maddelerin düzenlendikten sonra kullanılmasının

uygun olduğuna karar verilmiştir. 30 soruluk teste ait maddelerin büyük bir çoğunluğunun ayırt edicilik indeksinin 0.40'ın üzerinde olduğu görülmektedir. Yapılan düzenlemeler sonucunda testteki soru sayısı 29'a indirilmiş, yalnızca 3. soru testten çıkarılmış ve dört soru yeniden düzenlenerek teste dahil edilmiştir. 29 sorudan oluşan nihai başarı testindeki en düşük madde güçlük indeksi değerine (0.40) sahip olan madde 10., en yüksek madde güçlük indeksine (0.74) sahip olan madde ise 13. maddedir. Başarı testindeki en düşük ayırt edicilik indeksine (0.20) sahip olan maddeler 15. ve 19. maddeler iken, en yüksek ayırt edicilik indeksine (0.78) sahip olan maddelerin 8. ve 9. maddeler olduğu görülmüştür.

Güvenirlilik analizi. Akademik başarı testi puanları için hesaplanan ortalama (17.92), medyan (18.50) ve mod (18.00) değerlerinin birbirine yakın olması, çarpıklık (-0.455) ve basıklık (-0.444) değerlerinin -1, +1 aralığında olması nedeni ile verilerin normal dağıldığı söylenebilir. Çarpıklık katsayısının sıfıra oldukça yakın ve negatif değerde olması, dağılımın hafif sola çarpık olduğunu ve öğrencilerin başarı testinden aldıkları toplam puanların, ortalama olan 17.92'den daha büyük değerlerde toplandığını göstermektedir. Basıklık katsayısının da sıfıra yakın bir değer alması, dağılımın normale yakın olduğunu göstermektedir. Ayrıca öğrencilerin testte bulunan sorulardan ortalama %62'sini doğru cevaplandıkları saptanmış, toplam başarı testi puanlarına ilişkin standart sapma değeri 5.563 olarak hesaplanmıştır. Bu durum, öğrencilerin konuya ilişkin bilgi düzeyleri arasında büyük bir farklılık olmadığını göstermektedir.

Madde analizi sonucunda 29 maddeye indirgenmiş olan başarı testinin kullanımının geçerli olduğuna karar verilmiştir. Testin kullanımının güvenilir olup olmadığının belirlenmesi amacı ile KR-20 güvenilirlik testi uygulanmıştır. KR-20 yöntemine göre başarı testinin güvenilirliği 0.81 olarak hesaplanmış ve test güvenilir bulunmuştur (Cronbach, 1951).

Testin Cronbach Alpha güvenilirlik katsayısı 0.82 olarak bulunmuştur. Cronbach Alpha katsayısı ne kadar yüksek ise ölçek de o ölçüde birbiri ile tutarlı ve aynı özelliği yordayan maddelerden oluşmaktadır. Cronbach Alpha değeri 0.80 ile 1.00 aralığında olan ölçme araçları yüksek derecede güvenilir olarak kabul edildiğinden, geliştirilmiş olan başarı testinin yüksek güvenilirliğe sahip olduğu söylenebilir (Tavşancıl, 2006). Başarı testinin güvenilirlik analizi sonucunda, başarı testinden bir soru çıkarıldığında geride kalan soruların oluşturduğu test ortalamaları ve testin varyans değerleri incelenmiştir. Ayrıca ölçekten ilgili soru çıkartıldıktan sonra kalan bölüme ait güvenilirliği gösteren Cronbach Alpha değerleri de incelenmiştir. Buna göre, 1, 15, 19 ve 25. sorulara ilişkin düzeltilmiş madde-toplam korelasyonu değerleri diğer sorulara ait değerlerden daha düşüktür. Ancak bu sorular silindiğinde hesaplanan Cronbach Alpha değerlerine bakıldığında bu soruların çıkarılmasının testin Cronbach Alpha değerini değiştirmeyeceği görülmektedir. Dolayısı ile bu soruların tekrar gözden geçirilip düzenlenmesine karar verilmiştir.

Tablo 4

Akademik Başarı Testinin Son Haline İlişkin Betimsel İstatistikler

Öğrenci sayısı	Ortalama	Medyan	Mod	Standart Sapma	Çarpıklık	Basıklık
180	17.92	18.50	18.00	5.563	-0.455	-0.444

Son hali ile 29 maddeden oluşan akademik başarı testi puanları için hesaplanan ortalama (17.92), medyan (18.50) ve mod (18.00) değerlerinin birbirine yakın olması, çarpıklık (-0.455) ve basıklık (-0.444) değerlerinin -1, +1 aralığında olması nedeni ile verilerin normal dağıldığı söylenebilir (Tablo 4). Ayrıca öğrencilerin test sorularının ortalama %62'sini doğru cevaplandıkları saptanmış, toplam başarı testi puanlarına ilişkin standart sapma değeri 5.563 olarak hesaplanmıştır. Hesaplanan bu değer, öğrencilerin konuya ilişkin bilgi düzeylerinin birbirleri ile kıyaslandığında büyük bir farklılık göstermediğini belirtmektedir.

Faktör analizi. Nihai başarı testinin yapı geçerliliğinin sağlanması için, 0-1 şeklinde kategorilendirilen ikili veriler üzerinden hesaplanan ve çoktan seçmeli başarı testlerinde kullanılması önerilen (Uebersax, 2015) tetrakorik matrisin hesaplanmasına dayalı AFA yapılmıştır. AFA gerçekleştirilmeden önce veri yapısının faktörleşmeye uygunluğunun ve toplanan verilerin evreni temsil edip etmeme durumunun test edilmesi amacı ile Kaiser- Meyer-Olkin (KMO) ve Bartlett küresellik testi sonuçlarına bakılmıştır. Buna göre, testin KMO değeri 0.715, Bartlett's testi ise anlamlı ($p < .05$) bulunmuştur. Yani başarı testinin belirtilen amaç doğrultusunda kullanımı geçerli ve güvenilir olup, test faktör analizi için uygundur (Pallant, 2007). Yapılan AFA sonucunda, 29 maddenin, öz değeri 1'den büyük olan 11 faktör altında toplandığı ve bu faktörlerin, varyansın %66.29'unu açıkladığı belirlenmiştir. Bu yüzde, kabul edilebilir miktar olan %41'in oldukça üzerindedir (Kline, 1994). Çokluk, Şekercioğlu ve Büyüköztürk (2010), ölçek maddelerinin açıkladığı varyans miktarının %40 ile %60 arasında olmasının yeterli olduğunu belirtmektedir. Dolayısı ile ölçeğin yapı geçerliliğinin yüksek olduğu görülmektedir.

Analizler sonucunda elde edilmiş olan scree plot grafiğinde altıncı faktörden sonra ivmeli bir düşüş olduğu görülmüştür. Ayrıca, yapılan faktör analizinde ilk altı faktörün açıkladığı varyansın (%45.49) %41'den büyük olması ve sonraki faktörlerin açıkladıkları varyans ve özdeğerlerin azalması ölçeğin altı faktörlü bir yapıya sahip olabileceğini göstermektedir. Buna göre testin altı faktörden oluştuğu söylenebilir. Bu nedenle döndürmeye başvurularak tekrar AFA uygulanmıştır. Ölçekteki maddelerin faktör yük değerlerinin en az 0.30 olması tercih edilmektedir (Büyüköztürk, 2014). Tablo 5'te, tüm faktör yük değerlerinin 30'dan büyük olduğu görüldüğünden, testten çıkarılması gereken herhangi bir madde bulunmamıştır.

Tablo 5

Faktör Yük Değerleri

Madde	Faktör					
	1	2	3	4	5	6
1	.457					
4	.446					
6	.328					
8	.546					
10	.475					

Tablo 5

Devamı

Madde	Faktör					
	1	2	3	4	5	6
18	.706					
25	.648					
5		.971				
7		.387				
9		.455				
15		.721				
22		.747				
27		.519				
2			.381			
13			.535			
21			.334			
29			.312			
12				.776		
19				.395		
24				.541		
30				.420		
14					.322	
17					.639	
23					.509	
28					.378	
11						.495
16						.620
20						.357
26						.317

Başarı testinin scree plot grafiği ve faktör yük değerleri bir arada incelendiğinde altı faktörden oluştuğu görülmektedir. Bu nedenle bu altı faktör değerlendirmeye alınmıştır. Buna göre birinci faktörde yer alan maddeler: 1, 4, 6, 8, 10, 18 ve 25; ikinci faktörde yer alan maddeler: 5, 7, 9, 15, 22 ve 27; üçüncü faktörde yer alan maddeler: 2, 13, 21 ve 29; dördüncü faktörde yer alan maddeler: 12, 19, 24 ve 30; beşinci faktörde yer alan maddeler: 14, 17, 23 ve 28 ve altıncı faktörde yer alan maddeler: 11, 16, 20 ve 26. maddelerdir. Tablo 2’den yola çıkılarak her bir faktöre yerleşen maddeler incelendiğinde; birinci faktördeki 7 maddeden 5’inin öğretim programının “Aynalarda Yansıma ve Işığın Soğurulması” ünitesine ait birinci kazanıma ilişkin olarak hazırlanmış olduğu görülmektedir. Benzer şekilde ikinci faktörde bulunan 6 maddeden

4'ünün ikinci kazanıma, üçüncü faktörde bulunan maddelerin tümünün üçüncü kazanıma, dördüncü faktörde bulunan maddelerin tümünün dördüncü kazanıma, beşinci faktörde bulunan maddelerin tümünün beşinci kazanıma ve altıncı faktörde bulunan maddelerin tümünün altıncı kazanıma ait olduğu görülmektedir. Dolayısı ile tetrakorik faktör analizinde tespit edilen altı faktörlü yapı anlamlıdır.

Açık Uçlu Sorulara İlişkin Pilot Uygulama ve Analizler

Açık uçlu sorular üzerinde gerçekleştirilen pilot uygulama, verilen cevaplama süresinin yeterliğine karar verilmesi, soruların uygulama yönergelerinin yeniden düzenlenmesi, soruların okunabilirliği ve anlaşılabilirliğinin uygulama sırasında görülmesi, anlaşılması güç terimlerin belirlenerek değiştirilmesi, öğrenci düzeyine uygun olmayan maddelerin belirlenmesi ve rubrik geliştirilmesi amacıyla yapılmıştır. Hazırlanan soruların yaş grubuna uygun, açık ve anlaşılır olup olmadığının, verilen yanıtların sorulan soruların cevaplarını yansıtmadığının belirlenmesi amacı ile iç geçerliliğin sağlanabilmesi için sorular ikisi fen bilimleri, biri Türkçe eğitiminden üç uzman tarafından incelenmiş, uzmanların değerlendirmeleri sonucunda sorulara son hali verilmiştir. Soruların istenilen verileri sağladığı kanısına varılarak, soruların işlevliliğinin ölçülebilmesi amacı ile on öğrenci ile pilot çalışma yapılmıştır. Ayrıca ölçüğün okunurluk analizi, terimlerin anlaşılabilirlik analizi ve cümlelerin uzunluk analizleri yapılmıştır.

Açık uçlu maddelerden oluşan bir ölçme aracı kullanmanın en zayıf yönü değerlendirmedeki öznelliktir. Buradan yola çıkılarak, soruların iç tutarlılık güvenilirlik analizinin yapılması amacı ile Cronbach Alpha güvenilirlik katsayısı ve her bir soru için kdolayıcılar arası tutarlılık katsayısı hesaplanmıştır. Bunun için rastgele seçilen 10 katılımcının cevap kâğıtları puanlanmıştır. Açık uçlu soruların puanlanmasında araştırmacı tarafından geliştirilmiş olan cevap anahtarı ve dereceli puanlama anahtarı kullanılmıştır. Başarı testinin ortalama Cronbach Alpha güvenilirlik katsayısı 0.78 olarak hesaplanmıştır. Dolayısı ile açık uçlu maddeler güvenilirlerdir. Daha sonra, öğrencilerin cevapları ölçüm araçları kullanılarak, araştırmacı ve fen eğitiminde öğretim üyesi olan bir uzman tarafından puanlandırılmış ve kodlayıcılar-arası tutarlılığa bakılarak Kappa (κ) katsayısı hesaplanmıştır. Kodlayıcıların, aynı öğrencilerin aynı sorulara verdikleri cevaplar için vermiş olduğu kategorik puanlar Kappa testi yapılarak karşılaştırılmış ve uyuma oranları incelenmiştir. Bu çalışmada, Kappa uyuma katsayılarının ortalaması 0.82 olarak bulunmuştur. Bu da, puanlama araçlarının kodlayıcılar arası tutarlılık güvenilirliğinin mükemmel yakın olduğunu göstermektedir (Landis & Koch, 1977). Uzmanlar cevapları birbirlerinden bağımsız olarak ve rubrik kullanarak puanlamışlardır. Böylece uzmanların yedi soru için 10 öğrenciye verdikleri puanlardan oluşan bir veri seti elde edilmiştir. Açık uçlu sorulardan oluşan testten alınabilecek maksimum puan 100'dür. Puanlamalar arasındaki farkların karşılaştırılması için, hem her bir madde için ayrı ayrı, hem de tüm maddelerin toplam puanlarının ortalamaları arasındaki farklara ilişkin *t*-testi yapılmıştır (Tablo 6). Uzmanların her bir cevaba ve teste verdikleri toplam puanlar arasında anlamlı farklılık bulunamamıştır. Nitekim uzmanların toplam puan ortalamaları birbirine çok yakındır. Açık uçlu sorulara ait puanlama aracının farklı puanlayıcılar tarafından yüksek uyuma oranları ile aynı şekilde anlaşıldığı ortaya konulmuştur.

Tablo 6

Puanlayıcıların Ortalama Puanlarının Karşılaştırılması

Soru	1. Puanlayıcının ortalaması (\bar{P}_1)	2. Puanlayıcının ortalaması (\bar{P}_2)	Fark ($\bar{P}_1 - \bar{P}_2$)	t değeri	p
1	8.38	8.38	0.00	0.000	1.000
2	7.50	7.50	0.00	0.000	1.000
3	7.20	7.20	0.00	0.000	1.000
4	5.00	5.00	0.00	0.000	1.000
5	5.80	6.00	-0.20	-0.198	0.845
6	15.00	14.30	0.70	-0.151	0.881
7	14.05	13.35	0.70	0.308	0.762
Toplam	61.93	61.73	0.20	0.038	0.970

Sonuç ve Tartışma

Milli Eğitim Bakanlığı tarafından günün ihtiyaçları doğrultusunda değiştirilen öğretim programlarına paralel olarak, okullarda uygulanan ölçme ve değerlendirme araçlarının da yeniden ele alınarak güncellenmesi gerekmektedir. Bu amaçla 2015 7. sınıf öğretim programında yer alan “Aynalarda Yansıma ve Işığın Soğrulması” ünitesi ile ilgili çoktan seçmeli ve açık uçlu sorulardan oluşan bir başarı testi hazırlanarak geçerlilik ve güvenilirlik çalışması yapılmıştır. Başarı testine yönelik yapılan pilot uygulamalardan sonra yapılan güvenilirlik ve geçerlilik hesapları, geliştirilen testin belirtilen amaç için kullanımının güvenilir ve geçerli olduğunu ve ülkemiz genelindeki tüm 7. sınıf öğrencilerine, özellikle ünite sonunda öğrenme eksikliklerinin ölçülmesi ve değerlendirilmesi açısından uygulanabilir nitelikte bir başarı testi olduğunu göstermektedir. Bilindiği gibi, akademik başarı testlerinin ilgili konuda kapsam geçerliliğini taşıması en fazla önem taşıyan husustur (Büyüköztürk, 2014). Bu araştırmada, uzman görüşleri doğrultusunda yapılan iyileştirmelerle ilerlenen her aşamada daha tutarlı bir test geliştirildiği, dolayısıyla hazırlanan testin tutarlı olduğu söylenebilir.

Madde analizi sonucunda, uzman görüşü doğrultusunda oluşturulan başarı testinden bir soru çıkarılırken dört soru düzeltilmiştir. Buna göre testin güvenilirlik katsayısı 0.82 bulunmuştur. Öğrencilerin matematiksel, formüle ve ezbere dayalı soruları çözebilmelerinin kavramı anladıkları anlamına gelmediği bilinmektedir (Case & Fraser, 2001). Dolayısı ile başarı testinde bulunan soruların çoğu öğrencilerin alışageldikleri sayısal problemlerden daha çok kavramsal anlamaya yöneliktir. Dolayısı ile test sonuçları başarıya ait çıkarımlar yaparken aynı zamanda öğrencilerin olası kavram yanlışlarına veya öğrenme eksikliklerine de ışık tutabilir. Nitekim alan yazında da, başarı testlerinde kavram yanlışlarının çeldiriciler olarak kullanılmasının testin etkililiği açısından yararlı olduğunu öne süren bazı çalışmalar yer almaktadır (Çalık & Ayas, 2003; Treagust, 1988).

Özellikle eksik ve yanlış öğrenmelerin ortaya çıkarılmasında kullanılan çoktan seçmeli testler uygulama ve puanlama gibi pek çok avantaja sahip olmakla birlikte,

çeldiriciler sayesinde eksik ve yanlış öğrenmelerin belirlenmesini ve sonuçların genellenebilmesini sağlamaktadır (Ayvaci & Durmuş, 2016). Ancak bu testlerde sınırlı sayıda seçenek olduğundan öğrenciler fikirlerini özgürce ifade edememektedirler. Bunun yanı sıra öğrenciler bazen doğru cevabı bilmeden tesadüfi olarak da doğru seçeneği işaretleyebilmektedir (Mintzes, Wandersee, & Novak, 2000). Günümüzde bu dezavantajları en aza indirmek için, çoktan seçmeli sorulara ek olarak açık uçlu sorulara da yer veren başarı testleri geliştirilmeye başlanmıştır (örn. Günel, Uzoğlu, & Büyükkasap, 2009; Özmen & Kolomuç, 2004). Dolayısı ile bu çalışmada geliştirilmiş olan başarı testinde çoktan seçmeli sorulara ek olarak açık uçlu soruların da yer alması, ilgili testin literatürdeki önerileri karşılar nitelikte olduğu söylenebilir. Sonuç olarak geliştirilen başarı testi, hem fen bilimleri dersini yürüten öğretmenlere ölçme değerlendirme sürecinde geçerli ve yüksek güvenilirliğe sahip bir ölçme aracı, hem de fen eğitimi alanında bilimsel araştırma yapacak araştırmacıların öntest-sontest olarak kullanılabilecekleri bir ölçme aracının sağlanması yönü ile de önem taşımaktadır. 2018 yılında yenilenmiş olan fen bilimleri öğretim programında da aynı kazanımlar mevcut olduğundan, geliştirilmiş olan bu başarı testi güncelliğini korumaktadır. Dolayısı ile öğretmenler tarafından uygulanabilir.

Öneriler

Geliştirilmiş olan akademik başarı testi ile ilgili geçerlilik ve güvenilirlik çalışmalarından elde edilen bulgular göz önüne alındığında, öğretmenlere ve gelecekte yapılacak olan benzer çalışmalara ilişkin öneriler şunlardır:

1. Başarı testinin uygulanacağı örneklem sayısı daha fazla tutulabilir.
2. Farklı eğitim seviyelerinde bulunan öğrenciler seçilebilir. Böylece çalışma amacına daha iyi bir şekilde ulaşacaktır.
3. Test parametreleri kabul edilebilir düzeyde bulunmuş olsa da, madde analizi sonuçlarına göre başarı testi orta ve yüksek derecede başarıya sahip öğrencileri daha iyi ayırt etmektedir. Dolayısı ile test, düşük düzeyde başarıya sahip öğrencilerde iyi sonuçlar vermeyebilir. Test sonuçlarının öğrencilerin başarı seviyesine göre incelenmesi için başka çalışmalar yapılabilir.
4. Geliştirilmiş olan bu başarı testinin kullanılacağı yeni araştırmaların yapılması testin ölçme gücüne önemli katkılar sağlayacaktır.
5. Öğretmenler geliştirilmiş olan bu testi ünite öncesinde uygulayabilirler. Böylece test sonuçlarına göre öğrencilerin anlamakta güçlük çektiği kavramlar tespit edilerek öğretim süreci bu doğrultuda planlanabilir.

Summary

Purpose and Significance: For many years, Turkish students' science scores in international tests like PISA and TIMSS have been below the average. This situation shows that science education in our country could not be achieved at the desired level yet. Physics has a very important place in science. The conducted studies indicated that students are less successful in physics concepts of science rather than chemistry and biology concepts, and have negative attitude towards physics and have difficulty in learning physics (Bakaç & Kumru 1998). It is seen that even high school and university students have difficulty in explaining daily events such as light and image (Galili & Hazan, 2000; Kocakulah & Şardağ, 2013), and middle school students have many misconceptions about "Light and Sound" unit (Akdeniz, Yıldız, & Yiğit, 2001; Şen, 2003). If this topic is not fully understood in middle school, it will also be the basis for similar misconceptions in higher education. Whether a student is successful or not, or to what extent s/he is successful depend on assessment and evaluation process. In order to perform a proper assessment and evaluation, valid and reliable instruments are needed (Gönen, Kocakaya, & Kocakaya, 2011). Within this context, this study aims to develop a valid and reliable instrument to reveal the achievement of 7th grade students on "Light reflection in mirrors and light absorption" unit by considering the test development steps.

Method: Survey design was used in the study. The 37-item achievement test, consisting of a 30 multiple-choice and 7 open-ended questions, was developed and applied to 180 7th grade students from two different middle schools located in Kayseri province during spring semester of 2016-2017 academic year. In the tests development process, firstly, the course objectives to be measured in the test were determined. Then the educational objectives were determined by using the taxonomy created by Marzano and Kendall (2007). Expert opinions were obtained for the content validity and table of specifications was prepared. The multiple choice items were analyzed through MS Excel 2013, ITEMAN 4.2.1, SPSS 24 and FACTOR 10.8.04 programs, whereas open-ended items were analyzed using the Kappa (κ) coefficient.

Results: Item discrimination and difficulty indices were calculated using Excel and SPSS programs. The difficulty indices of the all items were found between 0.20-0.90. Since the average difficulty index of the items was found as 0.59, the test can be accepted as reliable (Hotiu, 2006). It is seen that the test was generally composed of easy and moderately difficult questions. The discrimination index of the third question was found as -0.20. This negative value indicates that this item does not serve the test purpose and that it does not distinguish between successful and unsuccessful students (Kubiszyn & Borich, 2003). Therefore, it was decided to remove this item from the test. Moreover, since the discrimination index of item 15 and 19 was found as 0.20; and item 1 and 25 was found as 0.29, it was decided to use these items after making some arrangements (Ebel & Frisbie, 1986). It is seen that the majority of the items have a discrimination index of over 0.40. The final form of the achievement test consisting of 29 questions had minimum and maximum item difficulty indices of 0.40 and 0.74

respectively, and minimum and maximum item discrimination indices of 0.20 and 0.78 respectively.

The Cronbach Alpha reliability coefficient of the test was found as 0.82. Also the reliability was calculated as 0.81 according to the KR-20 method. Hence, it can be said that the test has a high degree of reliability (Cronbach, 1951; Tavşancıl, 2006). Finally, exploratory factor analysis of a matrix of tetrachoric correlations was conducted to ensure that the final form of the test had construct validity. The scree plot graph showed an accelerated decline after the sixth factor. In addition to this, the factor analysis showed that the first six factors explain the variance which is greater than 41%, and that the variance and eigenvalues of the next factors are decreasing. Accordingly, it can be said that the achievement test is composed of six factors.

A pilot study was conducted with ten 7th grade students to measure the functionality of open-ended questions. The answer sheets of the students are scored using the rubric developed by the researchers. The reliability of open-ended questions was 0.78. Also the students' responses were scored by two experts and the Kappa (κ) coefficient was calculated. The mean of the Kappa coefficients of the open-ended questions was found as 0.82. Therefore, the open-ended questions are reliable (Landis & Koch; 1977). These results show that the achievement test, which consists of totally 36 questions, is a valid and reliable instrument that is applicable to all 7th grade students, especially at the end of the unit to measure and evaluate the learning deficiencies.

Discussion and Conclusions: In achievement tests, it is preferred that using open-ended and multiple choice questions together rather than to use one question type alone. Multiple-choice tests, especially used in the detection of incomplete and inaccurate learning, have many advantages such as the ease of implementation and scoring, and they enable generalization of the results (Ayvaci & Durmuş, 2016). However, since there are a limited number of options in these tests, students cannot express their opinions freely. In addition to this, students can sometimes randomly select the right option without knowing the correct answer. Therefore, using the open-ended questions as well as the multiple choice questions, it can be said that this achievement test can be considered to meet the recommendations in the literature.

This 36-item achievement test is a highly valid and reliable instrument that can be used by science teachers in the evaluation process of science courses, and it also can be used as pre-test and post-test by researchers who carry out scientific research in the field of science education.

Kaynakça

- Akdeniz, A. R., Yıldız, İ., & Yiğit, N. (2001). İlköğretim 6. sınıf öğrencilerinin ışık konusundaki kavram yanlışları. *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, 20(2), 72-78.
- Ayvacı, H. Ş., & Durmuş, A. (2016). Bir başarı testi geliştirme çalışması: Isı ve sıcaklık başarı testi geçerlik ve güvenilirlik araştırması. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 35(1), 87-102.
- Bahar, M., & Polat, M. (2007). The science topics perceived difficult by pupils at primary 6-8 classes: Diagnosing the problems and remedy suggestions. *Educational Sciences: Theory and Practice*, 7(3), 1113-1129.
- Bakaç, M., & Kumru, M. N. (1998). *Fen eğitiminde amaçların belirlenmesi*. III. Ulusal Fen Bilimleri Eğitimi Sempozyumu Bildirileri, 234-236, Trabzon.
- Boopathiraj, C., & Chellamani, K. (2013). Analysis of test items on difficulty level and discrimination index in the test for research in education. *International Journal of Social Science & Interdisciplinary Research*, 2(2), 189-193.
- Brooks, V. (2002). *Assessment in secondary schools: The new teacher's guide to monitoring, assessment, recording, reporting and accountability*. Buckingham: Open University Press.
- Büyüköztürk, Ş. (2014). *Sosyal bilimler için veri analizi el kitabı* (19. Baskı). Ankara: Pegem Akademi Yayıncılık.
- Case, M. J., & Fraser, M. D. (2001). An investigation into chemical engineering students' understanding of the mole and the use of concrete activities to promote conceptual change. *International Journal of Science Education*, 21(12), 1237-1249.
- Cochran Smith, M. (2001). Constructing outcomes in teacher education: Policy, practice and pitfalls. *Education Policy Analysis Archives*, 9(11), 1-68.
- Colin, P., & Viennot, L. (2001). Using two models in optics: Students' difficulties and suggestions for teaching. *American Journal of Physics*, 69(7), 36-44.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297- 334.
- Çalık, M., & Ayas, A. (2003). Çözümlerde kavram başarı testi hazırlama ve uygulama. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 14(14), 1-17.
- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları* (1. Baskı). Ankara: Pegem Akademi.
- Ebel, R. L., & Frisbie, D.A. (1986). *Essentials of educational measurement* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. (2012). *How to design and evaluate research in education* (8th ed.). New York: McGraw Hill.
- Galili, I., & Hazan, A. (2000). The influence of an historically oriented course on students' content knowledge in optics evaluated by means of facets-schemes analysis. *American Journal of Physics*, 68(S1), 3-15.

- Gronlund, N. E., & Linn, R. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Mac Millian Publishing.
- Gönen, S., Kocakaya, S., & Kocakaya, F. (2011). Dinamik konusunda geçerliliği ve güvenilirliği sağlanmış bir başarı testi geliştirme çalışması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 13(1), 40-57.
- Günel, M., Uzoğlu, M., & Büyükkasap, E. (2009). Öğrenme amaçlı yazma aktivitelerinin kullanımının ilköğretim seviyesinde kuvvet konusunu öğrenmeye etkisi. *Gazi Eğitim Fakültesi Dergisi*, 29(1), 379-399.
- Güneş-Koç, R. S. (2013). *5E modeli ile desteklenen bağlam temelli yaklaşımın yedinci sınıf öğrencilerinin ışık ünitesindeki başarılarına, bilgilerinin kalıcılığına ve fen dersine karşı olan tutumlarına etkisi* (Yayımlanmamış yüksek lisans tezi). Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Karamustafaoğlu, S., Bacanak, A., Değirmenci, S., & Karamustafaoğlu, O. (2010). Ses kavramına yönelik bir çoklu zeka etkinliği. *Batı Anadolu Eğitim Bilimleri Dergisi*, 1(2), 125-139.
- Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.
- Kocakülah, A., & Şardağ, M. (2013). Fen bilgisi öğretmen adaylarının görüntü oluşumu hakkındaki kavramsal anlamaları. *Eğitim ve Öğretim Araştırmaları Dergisi*, 2(4), 1-14.
- Kubiszyn, T., & Borich, G. (2003). *Educational testing and measurement: Classroom application and practice* (7th ed.). Hoboken, NJ: John Wiley & Sons, Inc.
- Kutu, H. (2011). *Yaşam temelli ARCS öğretim modeliyle 9. sınıf kimya dersi "Hayatımızda Kimya" ünitesinin öğretimi*. Yayımlanmamış doktora tezi, Atatürk Üniversitesi Eğitim Bilimleri Enstitüsü, Erzurum.
- Küçük, T. (2014). *Işık ünitesinde simülasyon yönteminin kullanılmasının öğrencilerin fen başarısına ve fen tutumlarına etkisi* (Yayımlanmamış yüksek lisans tezi). Çanakkale Onsekiz Mart Üniversitesi Eğitim Bilimleri Enstitüsü, Çanakkale.
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159-174.
- Marzano, R., & Kendall, J. (2007). *The new taxonomy of educational objectives* (2nd ed.). Thousand Oaks, CA: Corwin Press.
- MEB. (2007). *Ortaöğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı (OKS) 8. sınıf test kitapçığı*. Ankara: Eğitim Teknolojileri Genel Müdürlüğü, Ölçme ve Değerlendirme Dairesi Başkanlığı.
- MEB. (2009). *Seviye Belirleme Sınavı (SBS) 7. sınıf test kitapçığı*. Ankara: Eğitim Teknolojileri Genel Müdürlüğü, Ölçme ve Değerlendirme Dairesi Başkanlığı.
- MEB. (2011). *Seviye Belirleme Sınavı (SBS) 7. sınıf test kitapçığı*. Ankara: Eğitim Teknolojileri Genel Müdürlüğü, Ölçme ve Değerlendirme Dairesi Başkanlığı.
- MEB (2013). *PISA 2012 ulusal ön raporu*. Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü, Ankara. <http://pisa.meb.gov.tr/wp-content/uploads/2013/12/pisa2012-ulusal-on-raporu.pdf> adresinden 21 Ekim 2017 tarihinde edinilmiştir.
- MEB (2014). *TIMSS 2011 ulusal matematik ve fen raporu: 8. sınıflar*. Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü, Ankara. <http://timss.meb.gov.tr/wp->

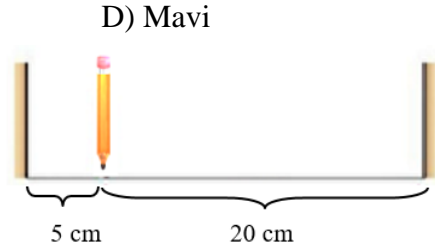
- content/uploads/TIMSS-2011-8-Sinif.pdf adresinden 21 Ekim 2017 tarihinde edinilmiştir.
- MEB (2015). *Fen bilimleri dersi 7. sınıf öğretim programı*. Ankara: Milli Eğitim Bakanlığı.
- MEB (2016a). *PISA 2015 ulusal raporu*. Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü, Ankara. http://odsgm.meb.gov.tr/test/analizler/docs/PISA/PISA2015_Ulusal_Rapor.pdf adresinden 21 Ekim 2017 tarihinde edinilmiştir.
- MEB (2016b). *TIMSS 2015 ulusal matematik ve fen bilimleri ön raporu: 4. ve 8. sınıflar*. Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü, Ankara. http://timss.meb.gov.tr/wp-content/uploads/TIMSS_2015_Ulusal_Rapor.pdf adresinden 21 Ekim 2017 tarihinde edinilmiştir.
- MEB (2018). *Fen bilimleri dersi 7. sınıf öğretim programı*. Ankara: Milli Eğitim Bakanlığı.
- Mertens, D. (2005). *Research methods in education and psychology: Integrating diversity with quantitative and qualitative approaches* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2001). Assessing understanding in biology. *Journal of Biological Education*, 35(3), 118-125.
- Nunnally, J. C. (1967). *Psychometric theory*. New York, NY: McGraw-Hill Book Company.
- Özmen, H., & Kolomuç, A. (2004). Bilgisayarlı öğretimin çözümler konusundaki öğrenci başarısına etkisi. *Kastamonu Eğitim Dergisi*, 12(1), 57-68.
- Özsevgeç, T., Çepni, S., & Bayri, N. (2007). Kalıcı kavramsal değişimde 5E modelinin etkililiği. *Yeditepe Üniversitesi Eğitim Fakültesi Dergisi*, 2(2), 36-48.
- Pallant, J. (2007). *SPSS survival manual: A step by step guide to data analysis using SPSS* (3rd ed.). Buckingham: Open University Press.
- Sayın, Ş. (2015). *İlköğretim fen ve teknoloji dersi 7. sınıf 'Işık' ünitesinin öğretiminde kavram karikatürleri kullanımının öğrencilerin akademik başarıları, sorgulayıcı öğrenme becerileri algıları ve motivasyonları üzerine etkisi* (Yayımlanmamış yüksek lisans tezi). Celal Bayar Üniversitesi, Fen Bilimleri Enstitüsü, Manisa.
- Şen, İ. A. (2003). İlköğretim öğrencilerinin ışık, görme ve aynalar konusundaki kavram yanlışlarının ve öğrenme zorluklarının incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25, 176-185.
- Tavşancıl, E. (2006). *Tutumların ölçülmesi ve SPSS ile veri analizi* (3. Baskı). Ankara: Nobel Yayın Dağıtım.
- TIMSS (the Trends in International Mathematics and Science Study) (2011). *Açıklanan fen soruları: 8. sınıf*. 28.11.2016 tarihinde, <http://timss.meb.gov.tr/> adresinden alınmıştır.
- 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.

- Uebersax, J. S. (2015). The tetrachoric and polychoric correlation coefficients. Statistical Methods for Rater Agreement. <http://john-uebersax.com/stat/tetra.htm> adresinden 20.01.2019 tarihinde erişilmiştir.
- Uzun, F. (2013). *Bağlam temelli yaklaşıma dayalı genel fizik laboratuvar dersinin fen bilgisi öğretmen adaylarının başarılarına, bilimsel süreç becerilerine, motivasyonlarına ve hatırlamalarına etkisi*. Yayınlanmamış yüksek lisans tezi, Marmara Üniversitesi Eğitim Bilimleri Enstitüsü.
- Uzunöz, A., & Buldan, İ. (2012). Ortaöğretim coğrafya dersi Doğal Sistemler konu alanı Atmosfer ve İklim ünitesi başarı testi geliştirme çalışması. *Kastamonu Eğitim Dergisi*, 20(1), 291-312.

EK-1

Akademik Başarı Testi – Çoktan Seçmeli Örnek Soru Maddeleri

6. Aşağıdaki araçlardan hangisinin çalışma prensibi, tümsek aynalara dayandırılmıştır?
A) Periskop B) Projeksiyon cihazı C) Tepegöz D) Dikiz aynası
12. Bir prizmaya beyaz ışık göndererek yaptığımız bir deneyde, ışık prizmadan çıkarken renklere ayrıldı. Hangi rengi göremezsiniz?
A) Sarı B) Turuncu C) Pembe D) Mavi
22. Bir kalem, birbirine paralel iki düz aynanın arasına şekildeki gibi konuluyor. Buna göre, kalemin aynalarda oluşan ilk görüntüleri arasındaki uzaklık kaç cm'dir?
A) 25 B) 50 C) 75 D) 100



EK-2

Akademik Başarı Testi – Açık Uçlu Örnek Soru Maddeleri

3. Şekildeki saatlerin düz aynadaki görüntülerine bakan Ayşe'nin saatleri kaç olarak göreceğini aynanın sağ tarafına çiziniz.

12:55



7. Sena Fen bilimleri öğretmenin verdiği araştırma ödevini yapmak için Güneş ışığını eşit açı ile alabilecekleri bir ortama beyaz, yeşil ve siyah renkli oyuncak arabaları koyuyor. Sıcaklıklarını ölçmek amacı ile arabaların içerisine özdeş termometreler yerleştiriyor. 10 dakika aralıklarla her bir araba için 3'er kez sıcaklık ölçümü yapıyor.
- a) Bu deneyin amacı ne olabilir?
b) Deneyin hipotezi ne olabilir?
c) Deneyin sonucunda ne gibi bir değişiklik meydana gelir? Bu durumun nedeni nedir?
d) Sena deney sırasında yapmış olduğu ölçümleri aşağıdaki grafiklere kaydetmek istiyor. Senanın çizeceği grafikler nasıl olmalıdır? Çizerek belirtiniz.



This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0). For further information, you can refer to <https://creativecommons.org/licenses/by-nc-sa/4.0/>