

8. SINIF ÖĞRENCİLERİNİN MEVSİMLERİN OLUŞUMU İLE İLGİLİ BİLGİLERİNİ ÖLÇMEYE YÖNELİK BİR BAŞARI TESTİNİN GELİŞTİRİLMESİ¹

DEVELOPMENT OF AN INSTRUMENT FOR MEASURING 8TH GRADE STUDENTS' KNOWLEDGE ABOUT FORMATION OF SEASONS

Aytekin BİRGİN*, Hasan ÖZCAN**

Geliş Tarihi: 29.10.2020
(Received)

Kabul Tarihi: 06.06.2022
(Accepted)

ÖZ: Bu çalışma ile 8. sınıf öğrencilerinin, mevsimlerin oluşumu konusundaki bilgilerini değerlendirebilmek için geçerli ve güvenilir bir başarı testinin geliştirilmesi amaçlanmaktadır. Mevsimlerin oluşumu konusu ile ilgili alanyazın incelendiğinde genellikle kavram yanlışları üzerine yapılan çalışmalar tespit edilmiş olup herhangi bir test geliştirme çalışmasına rastlanmamıştır. Çalışma tarama yöntemi ile yürütülmüş olup alanyazındaki başarı testi geliştirme basamakları kullanılmıştır. 2020–2021 eğitim öğretim yılı güz döneminde İç Anadolu Bölgesindeki bir il merkezi ile ona bağlı ilçe merkezlerindeki ortaokullarda öğrenim gören ve mevsimlerin oluşumu konusunu tamamlamış 347 sekizinci sınıf öğrencisi, araştırmanın örneklemini oluşturmaktadır. Çalışma kapsamında 2018 yılında yayımlanan fen bilimleri dersi öğretim programındaki F.8.1.1 Mevsimlerin Oluşumu konusuna yönelik 30 adet çoktan seçmeli sorunun yer aldığı bir başarı testi hazırlanmıştır. Testin geçerlik ve güvenilirlik çalışmaları yapılmıştır. Testin kapsam geçerliği için belirtke tablosu hazırlanırken yapı geçerliğini sağlamak için Faktör analizi yapılmış ve Kaiser-Meyer-Olkin (KMO) değeri 0.743 olarak bulunmuştur. Testin Kuder-Richardson 20 (KR-20) güvenilirlik katsayısı 0.83 olarak hesaplanmıştır. Yapılan analizler sonucu, düzenlenen veya çıkarılan maddeler olmakla birlikte teste kalması kararlaştırılan maddelerin ortalama güçlük ($p_j = 0.44$) ve ortalama ayırt edicilik ($r_j = 0.50$) değerlerinin uygun olduğu ifade edilebilir. Çalışma sonucunda mevsimlerin oluşumu konusunda 27 maddeden oluşan geçerli ve güvenilir bir başarı testi geliştirilmiştir.

Anahtar Kelimeler: Fen Bilimleri, Mevsimlerin Oluşumu, Başarı Testi, Test Geliştirme

ABSTRACT: This study aims to develop an achievement test that has measurement reliability and validity to evaluate the achievements of 8th-grade students about seasons.

¹Bu makale ilk yazarın, “STEM Yaklaşımı İle Mevsimlerin Oluşumu Konusunda Bir Öğretim Modülünün Geliştirilmesi Ve Ters Yüz Edilmiş Sınıflarda Uygulanabilirliğinin İncelenmesi” isimli yüksek lisans tez çalışmasından türetilmiş olup Aksaray Üniversitesi İnsan Araştırmaları Etik Kurulu’ndan 21.10.2020 tarih ve 2020/01-64 sayılı izin belgesi alınmıştır.

* Aksaray Ortaköy Fatih Ortaokulu, birginaytekin25@gmail.com, ORCID: 0000-0002-6894-6420.

** Doç. Dr., Aksaray Üniversitesi Eğitim Fakültesi, hozcan@aksaray.edu.tr, ORCID: 0000-0002-4210-7733

When the test on seasons was examined, it was seen that there was no test development work other than the studies on misconceptions. The survey method and test development steps are used within the scope of the study. The participants of the study consisted of 347 eighth-grade students who completed the subject of seasons and studied at schools in a province in Central Anatolia and its district centers in the 2020-2021 academic year. Within the scope of the study, an achievement test including 30 multiple-choice questions on F.8.1.1 Formation of Seasons in the science curriculum published in 2018 was prepared. The validity and reliability studies of the test have been done. While preparing the indicator table for the content validity of the test, FACTOR analysis was performed to ensure the construct validity and the Kaiser-Meyer-Olkin (KMO) value was found to be 0.743. The Kuder-Richardson 20 (KR-20) reliability coefficient of the test was calculated as 0.83. It was observed that the average difficulty ($p_j = 0.44$) and average discrimination ($r_j = 0.50$) values of the items that were decided to be in the test, although there were items that were edited or removed as a result of the analysis. As a result of the study, a valid and reliable achievement test consisting of 27 items was developed on the formation of the seasons.

Key Words: Science education, Formation of seasons, Achievement test, Test development

EXTENDED ABSTRACT

This section contains summary information about the purpose, importance, method, and results of the study.

Purpose

With this research, it is aimed to measure the understanding of the eighth-grade students about the formation of the seasons within the scope of the Science Curriculum. In this context, it is requested to be a test consisting of new generation questions prepared in different types and shapes, which can measure the meaningful learning of students towards their immediate environment, which is suitable for students' learning levels, and which prompts high-level thinking about abstract phenomena.

Importance

Although the subject of seasons is the first subject of the 8th-grade level today, it was left to the end of the academic year before 2018. Therefore, it could not find enough place in national exams. Parallel to this situation, it has been prepared within the framework of the Science Curriculum (FBDÖP) acquisitions in the literature; No success test has been found that comprehensively covers the causes, results, and investigations of the seasons in different geographical areas. It can be stated that studies on the formation of seasons are mostly aimed at detecting misconceptions (Öksüz & Demir, 2019; Rollins, Denton & Janke, 1983; Sadler, 1992; Sneider, Bar & Kavanagh, 2011). In the light of these studies, it is thought that an achievement test, which will be developed on the formation of seasons and presented to the use of practitioners, will contribute to the literature in terms of the application process and results.

Method

In this study, an achievement test was developed to measure the understanding of the eighth-grade students about the formation of the seasons, and validity and reliability analyzes were made. For the content validity of the test, a table of specifications was created within

the framework of expert opinion and KMO value was calculated. For construct validity, exploratory factor analysis was performed using a tetrachoric correlation matrix. The KR-20 coefficient was calculated to examine its reliability. During the data collection, 347 eighth-grade students were studied. In accordance with the test development study, an instant single scanning model was used. The instantaneous single scanning model is based on the instantaneous situation determinations to be made on the whole universe or on a sample or sample to be taken from the whole universe in order to reach a general conclusion about the universe consisting of many elements (Balcı, 2021; Bryman & Bell, 2015; Karasar, 1991).

Conclusion

In the study, an effective test consisting of new generation questions was tried to be developed due to the LGS exam system that changed in 2018. Since it was not sufficient for the questions in a test to only cover the acquisition or to exemplify the cognitive domain steps, the item type was tried to be diversified. Because different question types can be written covering the same acquisition and the same cognitive domain. For example, while one item measures the ability to interpret graphics together with the target achievement, another item measures attention besides the target achievement. Some of them are given in the abbreviated form in plain text, while others can be extended with images and preliminary information. In this sense, context-based questions are also included.

Construct validity was also tried to be ensured within the framework of the validity studies of the test. In this context, the tetrachoric correlation coefficient was calculated by performing factor analysis and it was seen that the construct validity of the test was ensured (see Table 7). It has been determined that each item in the measurement tool serves the purpose of the test and creates unity among themselves (Büyüköztürk, 2002).

In the development of MOBT, it was aimed to create the necessary awareness by focusing on the measurement reliability and informing the teachers and students in the schools selected for the sample. Thus, the KR-20 reliability data of the test were found to be quite satisfactory after the 12th question was removed ($\alpha=0.83$).

Since this achievement test, which was created on the subject of the seasons, was developed with the screening model, it can be stated that it is a scale that can be applied to 8th-grade students in terms of generalizability to the universe. It is thought that it will fill an important gap in this regard since it is a situation that is not encountered in the literature in terms of the subject chosen in the study. As a result of the research, a test consisting of 27 multiple-choice questions with validity (see Table 7) and measurement reliability (see Table 9) was developed in order to measure the understanding of secondary school eighth-grade students about the seasons (see Appendix A). In this context, it can be said that the test, which is prepared by considering the stages of classical test theory (Haladyna, 2004: 14, 15), can be used to measure the understanding of 8th-grade students about the seasons.

1. GİRİŞ

Ölçme, insanlık tarihi kadar eskilere dayanan bir kavram olmakla birlikte günlük hayatın her alanında kullanılmaktadır. Ölçme, gözlemlenen niteliklerin sayısallaştırılması işlemi olarak ifade edilebilir. Değerlendirme ise ölçme sonucunu birtakım kıstaslara bakarak bir karar verme sürecidir (Struyven, Dochy ve Janssens, 2005; Tekin, 1991). Değerlendirme olmadan ölçme işleminin tek başına pek bir

anlam ifade etmemesinden dolayı ölçme ve değerlendirme çoğu zaman bir arada kullanılırlar. Bu nedenle iki farklı kavram olmasına rağmen sık birbiri ile de karıştırılmaktadır. Ölçme ve değerlendirme gerek öğrencilerin hazırbulunuşluk düzeylerinin belirlenmesi gerekse öğrenme ve öğretme sürecine ilişkin hedeflerin gerçekleşme durumlarının tespit edilmesinde önemli bir rol oynamaktadır (Merenda, 2007). Bu nedenle etkili bir eğitim öğretim faaliyeti için ölçme ve değerlendirme sürecinin iyi planlanması gerekmektedir (Korkmaz ve Kaptan, 2005; Newfields, 2006).

Eğitim araştırmalarında hem ulusal hem de uluslararası büyük çaptaki sınavlar da dâhil olmak üzere başarı testleri, ölçme ve değerlendirme bağlamında kuşkusuz önemli bir yer tutmaktadır. Başarı testleri, öğrenci kazanımlarındaki değişimin gözlemlenmesi, analiz edilmesi ve değerlendirilmesine olanak tanıyarak anlamlı öğrenmelerin tespit edilmesine yardımcı olmaktadır (Berliner, 2020; Doğan, 2009; Gibson ve Mitchell, 2008; McMillan ve Schumacher, 2010; Tezbaşaran, 2008). Geçerliği ve güvenilirliği sağlanmış başarı testleri, öğrencilerin hedef kazanımlara ne derece sahip olduklarını göstermelerinin yanı sıra öğretim etkinliklerinin kalitesi ve süreç sonunda yapılabilecek değişiklikler konusunda yol gösterici bir görev üstlenirler (Aslan, Karamustafaoğlu ve Kurt, 2018). Başarı testleri öğrencilerin eğitim öğretim faaliyetleri öncesinde “hazırbulunuşluk testi” olarak, eğitim öğretim faaliyetleri devam ederken “izleme testi” olarak, eğitim öğretim faaliyetleri sonunda ise “erişi testi” olarak uygulanabilirler (Yıldırım, 1999). Çoğunlukla çoktan seçmeli türü tercih edilmekle birlikte açık uçlu, boşluk doldurmalı ve doğru-yanlış biçimlerinde de geliştirilebilen başarı testleri, öğrencilerin bir konu hakkındaki anlayışlarını belirlemek üzere kullanılırlar. Öğrenci başarısını değerlendirebilmek için geliştirilen her ölçme aracında olduğu gibi başarı testlerinin de artıları ve eksileri vardır. Örneğin açık uçlu sorulardan oluşan başarı testlerinde sınav başarısının sadece sorulara verilen cevapların kalitesi ile ilgili olmadığı, puanlayıcılar arasında not farklılıkları olabildiği hatta aynı puanlayıcının bir teste farklı zamanlarda vermiş olduğu notlar arasında da farklılıklar olabileceği bilinmektedir. Geleneksel sınavların öznelliğine yönelik bu eleştirinin, çoktan seçmeli sınavlara öncelik verilmesine neden olduğu ifade edilebilir (Levesque, 2011).

Çoktan seçmeli testler, objektif olarak puanlanabilmelerine rağmen, testin şekli ve test öğelerinin seçimi bakımından oldukça öznedir (Turgut ve Baykul, 2012). Bu öznelliğe rağmen çoktan seçmeli testler ile kalabalık gruplarda puanlama güvenilirliğinin sağlanması daha kolay olabilmektedir (Tekindal, 2017). Madde sayısının çok olması, çoktan seçmeli testlerdeki kapsam geçerliğini artıran bir etmendir (Güler, 2017). Çoktan seçmeli testlerde tahminle veya yanlış cevapları eleyerek doğru cevabı bulma ihtimali bulunmaktadır. Bu duruma karşı seçenek sayısı arttırılarak bir önlem alınabilir. Buna rağmen bunun kesin bir çözüm olduğu

söylenemez (Choi ve Park, 2016; Yazıcıoğlu, 2017). Çoktan seçmeli testler, öğrencilerin kendilerini ifade edebilme becerilerini ve yaratıcılıklarını sınırlasa da hazırlanacak soruların Bloom'un üst düzey basamakları olan analiz, değerlendirme ve üretme basamaklarına uygun olarak hazırlanması bu sınırlılığı azaltabilir (Küçükahmet, 2002). Bu bakımdan soruların Bloom Taksonomisi'ne uygun olarak hazırlanması, ölçme ve değerlendirmede önemlidir (Jiun, Kamarudin, Talib ve Hassan, 2018; Karamustafaoğlu ve Tutar, 2016). Ayrıca, Bloom (1956) Taksonomisi'ne göre hazırlanan test sorularının, dersin kazanımlarını yansıtması ve konunun amacına uygun geçerli ve güvenilir olması, soruların niteliğini artırmaktadır (Kızılkapan ve Bektaş, 2018; Pandra ve Mardapi, 2017). Nitelikli bir test geliştirmede ilk adımın, ilgili kazanımı tespit etmek olduğu ifade edilebilir. Sonraki süreçte, kazanıma uygun olarak hazırlanan sorular uzmanlar tarafından incelenir, geri dönütler sonrası sorular yeniden düzenlenir, madde analizleri yapılır ve testin nihai haline ulaşılabilir (Gönen, Kocakaya, Kocakaya, 2011; Güler, 2012; Öngören, Şahin ve Adalı, 2007; Tiruchittampalam, Nicholson, Levin ve Ferron, 2018). Bu yoldan farklı olarak başarı testleri, alanyazında mevcut bir testin çevirisi ile elde edilebileceği gibi (Akbulut ve Çepni, 2013) çeşitli sınavlara hazırlık kitapları, ders kitapları ve ulusal sınavlardaki sorular incelenerek de hazırlanabilmektedir (Uygur, 2009).

Alanyazında mevsimlerin oluşumu konusu ile ilgili geliştirilen başarı testi çalışmaları incelendiğinde, astronomi konularının içerisinde, birkaç soru ile ya da mevsimlerin oluşumu konusunda hazırlanmış açık uçlu veya kapalı uçlu sınırlı sayıdaki sorulardan oluşan çalışmalara rastlanmaktadır (Bolat ve Altınbaş, 2018). Açık uçlu sorular genellikle "Mevsimlerin nedeni nedir? Çizim yaparak açıklayınız"; çoktan seçmeli sorular ise genellikle "Mevsimlerin oluşumunun ana nedeni nedir", "Yazların, kış aylarından sıcak olmasının nedeni nedir?" ya da "Güney ve kuzey yarım kürede aynı anda farklı mevsimler yaşanıyor, bunun nedeni nedir?", "Bulduğunuz yerden bayrak direğinin gölgesi ne zaman sıfır olur" şeklindeki sorulardır. Bazı çoktan seçmeli sorular da açık uçlu soru şekline çevrilerek kullanılmaktadır (Alkış, 2006; Atwood ve Atwood, 1996; Frede 2006; Gülen, 2019; Kikas, 1998; Sadler, 1992; Schoon, 1992; Sharp, 1996; Trumper, 2001a, 2001b; Türk ve Kalkan, 2015; Türk, Kalkan, Kıroğlu ve İskeleli, 2016). Bu çalışmalardan birinde, mevsimler ve iklim ünitesi ile ilgili öğrenci başarısının tespit edilmesi amacıyla, 6 açık uçlu soru sorulmuş ve öğrenci kazanımları yorumlanmaya çalışılmıştır (Gülen, 2019). Alkış da (2006) çalışmasında, öğrencilerin, mevsimlerin oluşumuyla ilgili düşüncelerini tespit etmek amacıyla bir veri toplama aracı hazırlamıştır. Bolat ve Altınbaş (2018) ise farklı bölümlerde öğrenim gören eğitim fakültesi öğrencilerinin mevsimler konusuyla ilgili anlayışlarını paralel karma yöntemi ile belirlemeye çalışmışlardır. Yapılan bu çalışmalar kavram yanlışları bağlamında değerlendirildiğinde, öğrencilerin özellikle mevsimlerin oluşumu ile

ilgili Dünya'nın Güneş'e yaklaştıkça sıcaklığın arttığını, uzaklaştıkça da sıcaklığın azaldığını savunan, "uzaklık teorisine" sahip oldukları tespit edilmiştir. Bu nedenle de ölçme araçlarında uzaklık teorisi ile ilgili maddelere de yer verilmesi gerekliliği belirtilmiştir. Ayrıca ilgili çalışmalarda mevsimlerin nedenleri kadar sonuçları için de maddelere yer verilmesi gerektiği söylenmiştir.

Mevsimler konusu, günümüzde 8. sınıf düzeyinde öğretim yılının ilk konusu olarak işlenirken, 2018 öncesinde öğretim yılının sonuna bırakılmaktaydı. Bundan dolayı ulusal sınavlarda yeteri kadar kendine yer bulamamıştır. Bu duruma paralel olarak alanyazında Fen Bilimleri Dersi Öğretim Programı (FBDÖP) kazanımları çerçevesinde hazırlanmış, 8. sınıf düzeyinde, mevsimler konusunda; mevsimlerin nedeni, sonuçları ve farklı coğrafi alanlardaki incelemelerini kapsamlı olarak ele alan bir başarı testine rastlanmamıştır. Mevsimlerin oluşumu konusundaki çalışmaların daha ziyade kavram yanlışlarının tespit edilmesine yönelik olduğu ifade edilebilir (Öksüz ve Demir, 2019; Rollins, Denton ve Janke, 1983; Sadler, 1992; Sneider, Bar ve Kavanagh, 2011). Bu çalışmalar ışığında mevsimlerin oluşumu konusunda geliştirilerek uygulayıcıların kullanımına sunulacak olan bir başarı testinin, uygulama süreci ve sonuçları itibariyle alanyazına katkı sunacağı düşünülmektedir.

1.1. Araştırmanın Amacı

Bu araştırma ile sekizinci sınıf öğrencilerinin Fen Bilimleri Dersi Öğretim Programı kapsamında, mevsimlerin oluşumu konusuyula ilgili anlayışlarının ölçülmesi amaçlanmaktadır. Bu bağlamda geçerli ve güvenilir bir test geliştirilerek uygulayıcıların kullanımına sunulması hedeflenmektedir.

Bu doğrultuda aşağıdaki sorulara yanıtlar aranmıştır:

- 1) Sekizinci sınıf öğrencilerinin mevsimlerin oluşumu konusu ile ilgili anlayışlarını ölçmeye yönelik olarak geliştirilen test geçerli midir?
- 2) Sekizinci sınıf öğrencilerinin mevsimlerin oluşumu konusu ile ilgili anlayışlarını ölçmeye yönelik olarak geliştirilen test güvenilir midir?

2. YÖNTEM

2.1. Araştırma Modeli

Bu araştırmada sekizinci sınıf öğrencilerinin mevsimlerin oluşumu konusuna yönelik anlayışlarını ölçmek amacıyla bir başarı testi geliştirilmiş ve geçerlik, güvenilirlik incelemeleri yapılmıştır. Bu sebeple, test geliştirme çalışmasına uygun olarak anlık tekil tarama modeli kullanılmıştır. Anlık tekil tarama modeli çok sayıda elemandan oluşan evren hakkında genel bir kanıya varmak için, tüm evren üzerinde veya ondan alınacak bir grup örnek ya da örneklem üzerinde yapılacak olan anlık durum saptamalarına dayanır (Balcı, 2021; Bryman ve Bell, 2015; Karasar, 1991).

2.2. Evren ve Örneklem

2020-2021 eğitim öğretim yılında İç Anadolu Bölgesindeki bir il merkezi ile ona bağlı ilçe merkezlerindeki ortaokullarda öğrenim gören 8. sınıf öğrencileri bu

araştırmanın evrenini oluşturmaktadır. Araştırmanın örneklemini ise araştırma evreninde yer alan 7 devlet okulundaki, 8. sınıf düzeyinde Destekleme Yetiştirme Kursuna (DYK) düzenli devam eden ve mevsimler konusu öğrenimini tamamlamış 347 öğrenciden oluşmaktadır. Öğrencilerin okullara göre dağılımı Tablo 1’de verilmektedir.

Tablo 1: Araştırmanın Çalışma Grubunu Oluşturan Okullar ve Öğrenci Dağılımı

Okul	İL/İlçe	Öğrenci Sayısı	Yüzde (%)
A	İlçe	25	7.2
B	İlçe	40	11.5
C	İlçe	27	7.7
D	İlçe	16	4.6
E	İl Merkezi	132	38
F	İl Merkezi	47	13.5
G	İl Merkezi	60	17.2
Toplam		347	100

2.3. Mevsimlerin Oluşumu Konusuna Yönelik Başarı Testinin (MOBT) Geliştirme Süreci

Başarı testinin geliştirilme sürecinde Haladyna’nın (2004: 14, 15) test geliştirme basamakları izlenmiştir.

2.3.1. Testin Amacının Belirlenmesi

Geliştirilen başarı testi, hâlihazırda yürürlükte olan FBDÖP’de yer alan 8. sınıf mevsimlerin oluşumu konusundaki kazanıma yönelik olarak öğrencilerin öğrenme düzeylerine uygun ve uygulayıcı öğretmenlerin faydalanması amacıyla hazırlanmıştır. Ayrıca geliştirilen testin, öğrencilerin yakın çevresine yönelik anlamlı öğrenmelerini ölçebilen, soyut olgulara yönelik üst düzey düşünmeye sevk eden, farklı tip ve şekillerle hazırlanmış yeni nesil sorulardan oluşan bir test olması istenmektedir.

2.3.2. Kapsamın Belirlenmesi

Bu bölümde, testin kapsamı belirlenirken 2018 yılında yayımlanan FBDÖP’de yer alan ilgili ünite ile ilgili bilgilerden yararlanılmıştır (MEB, 2018). Mevsimler konusunda yer alan genel bilgiler Tablo 2’de paylaşılmaktadır.

Tablo 2: Mevsimlerin Oluşumu Konusuna Ait Genel Bilgiler, Kazanım ve Uyarılar

Ünite Adı	Konu Alanı	Konu	Kazanım	Ders Saati	Uyarılar
Mevsimler ve İklim	Dünya ve Evren	Mevsimlerin Oluşumu	F.8.1.1.1. Mevsimlerin oluşumuna yönelik tahminlerde bulunur.	8	a. Dünya’nın dönme eksenine olduğuna değinilir. b. Dünya’nın dönme eksenine ile Güneş etrafındaki dolanma düzlemi arasındaki ilişkiye değinilir. c. Işığın birim yüzeye düşen enerji miktarının mevsimler üzerindeki etkisine değinilir.

Tablo 2 incelendiğinde, Mevsimlerin Oluşumu Konusunun 1 kazanımdan oluşması ve önerilen sürenin 8 ders saati olması, içeriğinin az olduğu izlenimini vermektedir. Buna rağmen konunun soyut kavramlar içermesi ve kavram

yanılgılarının fazlaca yaşanması konunun detaylıca incelenmesi gerekliliğini ortaya koymaktadır (Türk, Alemdar ve Kalkan, 2012).

Mevsimlerin oluşumu ile ilgili, öğretim programının açıklamalar ve uyarılar kısmı ile ders kitabındaki konu başlıkları kısmında yeteri kadar açıklayıcı bilgi verilmediği görülmüştür. Bir tek kazanımı dikkate almak, konu kapsamını eksik örnekleyebileceği gibi aynı konu alanındaki benzeşen soruların gereksiz sayıda yazılmasına da sebep olabileceği düşünülmüştür. Bu sebeple alanyazın taraması yapılarak mevsimlerin nedenleri ve sonuçlarının birlikte araştırıldığı çalışmalar (Alkış, 2006; Bolat ve Altınbaş, 2018; Gülen, 2019; Sneider et al., 2011) incelenmiştir. Buradan elde edilen veriler ışığında LGS örnek soruları analiz edilmiş ve mevsimler konusu için soru grupları oluşturulmuştur.

Soruların taksonomiye uygun olarak hazırlanması dersin amacına ve kapsamına uygun, geçerli ve güvenilir olarak hazırlandığını göstermektedir (Kızılkapan ve Bektaş, 2018). Bu nedenle Bloom (1956) tarafından geliştirilen ve daha sonra Haladyna'nın (1997) yüksek düzeyli düşünme becerileri bağlamında güncellediği tek boyutlu bilişsel alan taksonomisi kullanılmıştır. Bu sınıflamada anlama, problem çözme, yaratıcılık ve eleştirel düşünme basamakları bulunmaktadır. Çoktan seçmeli sorular yaratıcılık basamağına uygun olmadığı için bu basamak kullanılmamıştır. Programda belirtilen kazanıma ait soru grupları, soru numaraları ve Haladyna (1997) Taksonomisi ile ilgili sınıflamaya ait belirtke tablosu Tablo 3'te sunulmaktadır.

Tablo 3: MOBT 'ye Ait Soru Grupları ve Belirtke Tablosu

Kazanım	Kazanıma Ait Soru Grupları	Bilişsel Özellikler		
		Anlama	Problem Çözme	Eleştirel Düşünme
F.8.1.1.1. Mevsimlerin oluşumuna yönelik tahminlerde bulunur.				
Mevsimlerin Nedeni	Eksen Eğikliği	22		
	Dolanma Hareketi	18		
	Güneş'e Yakınlık		9	1
Mevsimlerin Oluşum Sürecindeki Değişimler	Güneş Işınlarnın Gelme Açısındaki Değişimler			4, 6
	Sıcaklıktaki Değişimler			17, 26
	Birim Yüzeye Aktarılan Enerji Miktarındaki Değişimler			25
	Gölge Boyundaki Değişimler	2		5
	Eşit Miktardaki Işınlarnın Etki Ettiği Aydınlanma Alanındaki Değişimler			20, 28
	Gece-Gündüz Sürelerindeki Değişimler			8, 15
Coğrafik Konuma Bağlı Olarak Mevsimler	Türkiye üzerindeki konumlara etkileri ve karşılaştırmalar	23	16	13
	Dünya üzerindeki konumlara etkileri ve karşılaştırmalar	7, 27	30	10, 11, 12, 24, 29
	Dünya ve Güneş'in konumları itibariyle oluşan karşılaştırmalar	3, 19, 21		14
Toplam %		30	10	60

Tablo 3'te verilen bilgiler ışığında çalışmada belirtilen hususlara dikkat edilerek testin kapsamı belirlenmiştir. Madde seçimi sırasında bilimsel süreç becerilerini dikkate alarak, alan bilgisi ile beraber tahmin (madde;13,25) ve çıkarım yapmaya dayalı (madde;10,16,27,28,30) çeşitli becerileri ölçen sorular da seçilmeye çalışılmıştır.

2.3.3. Soru Tipi, Sayısı ve Süresi

Başarı testinde sorular çoktan seçmeli ve dört seçenekli olarak hazırlanmıştır. Soruların seçiminde öncelikle 2018 yılında yayımlanan FBDÖP'deki kazanımlar ve kazanımların uyarıları, detaylı olarak incelenmiştir. Soruların hazırlanma ve geliştirilme sürecinde 8. sınıf ders kitabı (Yancı, 2020), konu hakkında yerli ve yabancı araştırmalar (Akbulut ve Çepni, 2013; Bolat ve Altınbaş, 2018; Sneider et al., 2011; Sontay ve Karamustafaoglu, 2020; Türk et al., 2012) detaylı bir şekilde analiz edilmiştir. Test geliştirilirken, çoktan seçmeli soru hazırlamada temel ilkelere uygun olarak Milli Eğitim Bakanlığı tarafından önceki yıllarda merkezi olarak uygulanmış Liselere Geçiş Sınavı (LGS), Ölçme Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü (ODSGM) örnek sınav soruları (Uygun, 2009) incelenerek hata ihtimalinin düşük olduğu düşünülen sorular kullanılmıştır (Özcan, Koca ve Söğüt, 2019). Bu sorular belirtilen kaynaklardan değişiklik yapılmadan doğrudan alınmıştır.

Levesque (2011) test sonuçlarını büyük ölçüde etkileyen süreci eleştirerek genellikle "güvenirlik" olarak adlandırılan, test sorularını örnekleyebilmenin yeterli olmadığını ifade ederek beş, on veya yirmi soruluk tipik bir sınavın, öğrenci yanıtlarına güvenilir bir temel oluşturmayacağını belirtmiştir. Bu eleştiri nedeniyle var olan kazanımdan soru çeşitlendirilmiştir. Soruların cevaplanma süresinin bir ders saatini geçmemesi istenildiğinden 30 maddelik bir test amaçlanmıştır. Madde seçiminin sağlıklı olmama ihtimali, uzman görüşüyle elenebilir olması veya belirtke tablosunda istenen dağılımın oluşmama durumu gibi nedenlerden dolayı başlangıç olarak 40 maddelik bir soru havuzu hazırlanmıştır. Demirkuş'un da (2019) altını çizdiği şekilde bireyin okulda öğrendiklerini günlük yaşamına aktarabilme yeteneği, başarının bir ölçütü olduğu için hazırlanan soru havuzuna günlük yaşamdan ve yakın çevreden örnek sorular da eklenmiştir. Ön pilot uygulama öncesinde soru sayısı Haladyna (1997) Taksonomisinin düşünme süreçleri, Tablo 3'teki soru grupları ve uzman görüşü de göz önüne alınarak amaçlandığı gibi 30 olarak belirlenmiştir. Teste dahil edilmeyen 10 madde seçilirken benzer konuya yönelik, benzer şekil ve çizime sahip, aynı bilişsel alanda yoğunlaşan maddeler tercih edilmiştir. Testte yer verilen sorulara ilişkin bilgiler Tablo 4'de sunulmaktadır.

Tablo 4: MOBT Soru Maddelerinin Alındıkları Sınavlar

Soru Kaynakları	Soru Numaraları	Yayımlanma Yılı
LGS Hazırlık Örnek Soruları	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15,16,19, 20, 22, 23, 26, 27, 29,30 (Testten çıkarılan 10 soru)	2019-2020
Kazanım Kavrama Testleri	7, 17, 18, 21, 24, 28	2019-2020
LGS	25	2020

2.3.4. Testin Geçerliği

Başarı testinin geliştirilmesinde ilk adım kapsam geçerliğini sağlamaktır. Kapsam geçerliğini artıran etkenlerden birisi de belirtke tablosu hazırlamaktır (DiDonato-Barnes, Fives ve Krause, 2014; Lambie, Blount ve Mullen, 2017). Bu çalışmada da bu nedenle belirtke tablosu hazırlanmıştır (Tablo 3). Soru grupları ile düşünce süreçlerinin sınıflandırılması amacıyla iki ölçme değerlendirme uzmanı öğretim üyesinin, iki de fen eğitimi alanında uzman öğretim üyelerinin görüşüne başvurulmuştur. Öğretim üyeleri Tablo 3'te bulunan soru gruplarını dikkate alarak hazırlanan maddelerin, hedef kazanımı ölçülebilir nitelikte olduğunu ifade etmişlerdir. Uzman görüşlerinden gelen öneriler ve düzeltmeler neticesinde başarı testine son hali verilmiştir (bkz. Tablo 3). Bu değişiklikler sonucu ön pilot uygulama öncesi toplam 30 sorudan oluşan başarı testi hazır hale getirilmiştir.

Testin yapı geçerliği sonuçları da incelenmiştir. Yapı geçerliği, teste ölçülmesi istenen davranışlar açısından soyut kavramları doğru olarak ölçebilme derecesini göstermektedir (Ergin, 1995). Puanlaması 1-0 şeklinde olan testlerde yapı geçerliğinin belirlenmesi için tetrakorik korelasyon matrisi kullanılarak açılımlayıcı faktör analizinin yapılması uygun görülmektedir (Büyüköztürk, 2002). Testin kapsam geçerliğinin uygun olduğuna karar vermek için de KMO değerinin 0.70 değerinin üzerinde bulunması gerekmektedir (Baran, 2018).

2.3.5. Ön Pilot Uygulama

Pilot çalışmalar bazen "öncü denemeler" olarak adlandırılır ve büyük ölçekli araştırmaların planlanmasına rehberlik ederler. Ön pilot uygulamalar, çalışmanın amacına bağlı olarak değişmekle birlikte pilot uygulamadaki örneklem büyüklüklerinden küçük olarak planlanır (Yin, 2009). Bu bağlamda örneklem büyüklüğü belirlenirken, okulların salgın sürecinde açılmamış olmasından dolayı DYK'ya katılan az sayıdaki öğrenci varlığı göz önüne alınmıştır. Pilot uygulama öncesi, testin uygulanması sırasında cevaplama süresinin yeterli olup olmadığının ve soruların öğrenciler tarafından anlaşılma durumunun tespiti için 34 sekizinci sınıf öğrencisiyle ön pilot uygulama gerçekleştirilmiş; öğrencilerden gelen geri bildirimler neticesinde, 10. ve 12. maddelerde görseller üzerindeki yazılar netleştirilmiş ve büyük yazılmıştır. Bununla birlikte testten çıkarılan veya düzeltilen sorular ile ilgili madde analizleri Tablo 5'te paylaşılmaktadır.

Tablo 5: Soru Maddelerine Ait Madde Güçlük İndeksi ve Ayırt Edicilik

Soru Numaraları	Madde Ayırt Edicilik İndeksi	Madde Güçlük İndeksi
Ön Pilot 1. soru	0.11	0.94
Ön Pilot 14. soru	0.11	0.88
Ön Pilot 8. soru	0.22	0.76
Ön Pilot 13. soru	-0.33	0.09

Tablo 5 incelendiğinde 1. ve 14. maddeler, kolay olmalarından dolayı madde ayırt edicilikleri çok düşük değerler almıştır. Bu nedenle bu sorular testten çıkarılmıştır. 8. soru üzerinde düzenleme yapılarak teste dâhil edilmiştir. 13. soru özelinde öğrencilerle görüşme sonrası yapılan analizde sorudaki grafiğin yanlışlığına sebep olduğu fark edilmiştir. Sorunun kazanımlar açısından önemli olması nedeniyle sorudan grafik kısmı çıkarılarak uzman görüşleri ile desteklenerek yorum yazısı eklenmiştir.

2.3.6. Pilot Uygulama

Ön pilot uygulama sonucu 28 soruya düşürülen MOBТ, yedi farklı okulda bulunan toplam 384 öğrenciye uygulanmıştır. Uygulama sonucu hatalı, eksik ya da özensiz olduğu fark edilen sorular çıkarılmış ve 347 öğrencinin test sonuçları analize dâhil edilmiştir. Alanyazın incelendiğinde uygulamaya katılan öğrenci sayısının belirlenmesinde birtakım kriterler olduğu görülmüştür. Bunlar öğrenci sayısının, testteki madde sayısının 5 katı (Tavşancıl, 2010) ya da 10 katı olması veya öğrenci sayısının 200’den az olmaması (Andrew, Pedersen ve McEvoy, 2011; Kline, 2005) yönündeki görüşlerdir. Bu görüşler doğrultusunda bu araştırmanın örneklem büyüklüğünün yeterli olduğunu söylemek mümkündür.

2.3.7. Güvenirlik ve Madde Analizi

Öğrencilerin bir testten aldığı puanların güvenirliliğini belirlemede çeşitli istatistikî yöntemler bulunmaktadır. Bu çalışmada öğrenci puanları normal dağılım gösterdiği için güvenirlilik analizinde sıklıkla tercih edilen yöntemler arasında yer alan Kuder-Richardson 20 (KR-20) iç tutarlılık kat sayısı hesaplanarak yorumlanmıştır. Çalışmadan elde edilen KR-20 kat sayısı ($0.7 \leq \alpha < 0.9$) uygun bir değerdir (George ve Mallery, 2003). Testte, maddelerin doğru yanıtlanma durumlarını ortaya koymak için “madde güçlük indeksi” ve maddelerin neyi ölçmesi isteniyorsa o özelliği taşıyıp taşımadığının ayırt edebilirliğini tespit etmek için de “madde ayırt edicilik indeksi” hesaplanmıştır (Hasançebi, Terzi ve Küçük, 2020). Öğrencilerin başarı testinden aldıkları puanlar, madde analizleri yapmak amacıyla yukarıdan aşağıya doğru sıralanmıştır. Burada %27’lik ilk bölüm “üst grup”, son %27’lik bölüm ise “alt grup” şeklinde tanımlanmıştır (Khairani ve Shamsuddin, 2016). Çalışmadan elde edilen madde ayırt edicilik indeksi ve madde güçlük indeksi alanyazına dayalı olarak Tablo 6’da verilen kıstaslara göre yorumlanmıştır (Karşlı ve Ayas, 2013).

Tablo 6: Madde Ayırt Edicilik İndeksi ve Madde Güçlük İndeksi Kıstasları

Madde Ayırt Edicilik İndeksi	0.40 ve daha üstü	Çok iyi madde
	0.30 ile 0.39 arası	Oldukça iyi madde
	0.20 ile 0.29 arası	Düzenlenip, geliştirilmeli
	0.19 ve daha düşük	Çok zayıf madde, testten çıkarılmalı
Madde Güçlük İndeksi	0.61 ve yukarısı	Kolay madde
	0.60-0.40	Orta güçlükte madde
	0.39 ve aşağısı	Zor madde

3. BULGULAR

Bu bölümde MOB'T'ye ait geçerlilik, madde analizi ve ölçüm güvenirliği çalışmalarına yer verilmiştir.

3.1. MOB'T'nin Faktör Yapı Geçerliliğine Ait Bulgular

Testin faktör analizine uygun olup olmadığını anlamak için öncelikle küresellik testi yapılmış ve Bartlett Testi P değeri 0.000010 olarak anlamlı ($p < 0.05$) bulunmuştur. Yapılan analiz sonucu MOB'T'nin faktör yapı geçerliliğine ait bulgular Tablo 7'de sunulmuştur.

Tablo 7: MOB'T'nin Faktör Yükleri ve Açıkladıkları Varyans Oranları

Bağımlı Değişken	Faktör Yüğü	Ortak Varyans
Madde 1	Çıkarılan 1	
Madde 2	0.361	0.130
Madde 3	0.516	0.267
Madde 4	0.668	0.446
Madde 5	0.524	0.275
Madde 6	0.490	0.240
Madde 7	0.422	0.178
Madde 8	0.309	0.096
Madde 9	*0.279	0.107
Madde 10	0.575	0.330
Madde 11	0.606	0.367
Madde 12	*0.047	0.002
Madde 13	0.571	0.326
Madde 14	Çıkarılan 2	
Madde 15	0.432	0.187
Madde 16	0.634	0.401
Madde 17	0.598	0.358
Madde 18	*0.294	0.115
Madde 19	0.672	0.451
Madde 20	0.520	0.270
Madde 21	0.722	0.521
Madde 22	0.620	0.384
Madde 23	*0.283	0.111
Madde 24	0.581	0.337
Madde 25	0.604	0.364
Madde 26	0.634	0.402
Madde 27	0.702	0.493
Madde 28	0.337	0.114
Madde 29	0.500	0.250
Madde 30	0.435	0.189

*: Ölçekten çıkarılması gereken madde (< 0.30)

Tablo 7'de bulunan maddelerin faktör yükleri incelendiğinde 11. maddenin faktör yükü çok düşük bulunmuştur. Diğer 9, 18, ve 23. maddelerinin faktör yükü 0.30 değerinin altında kalmasına rağmen sınır değere yakın oldukları görülmüştür. Kapsam geçerliğini düşürmemek için bu maddeler testten çıkarılmamışlardır. Verilerin faktör analitik modeli ile modellenip modellenmeyeceğinin anlaşılması için faktör analizi 12. madde çıkarılarak tekrar yapılmış ve KMO değeri 0.743 olarak bulunmuştur. Bu değer 0.70 üzerinde olduğu için testin KMO değeri orta düzeyde

geçerli bulunmuştur. Ölçek geliştirme sürecinde seçilen örneklem büyüklüğünün analiz için yeterli olduğunu söylenebilir.

Goodness of Fit Index (GFI) değeri de 0 ile 1 arasında ve 0.941 olarak bulunmuştur. GFI değerinin 0.90 üzerinde yer alması mükemmel bir model göstergesi olarak kullanılmaktadır. Ayrıca gözlenen değişkenler arasında yeterli kovaryansın hesaplandığı anlaşılmaktadır.

3.2. MOBT 'nin Madde Analizine Ait Bulgular

Teste eklenen soruların yapı geçerliliğinin sağlanması için madde analizi yapılması sonucu hesaplanan değerler Tablo 6 dikkate alınarak düzenlenmiş ve Tablo 8'de verilmiştir.

Tablo 8: MOBT Madde Analizi

Soru No	Üst Grup (%27=100) Alt Grup (%27=100)	Doğru Cevap	Ayrıt Edicilik İndeksi (D)	Ayrıt Edicilik Değerlendirme	Güçlük İndeksi (P)	Güçlük Değerlendirme
1	Çıkarılan 1					
2	Üst grup	78	0.36	Oldukça İyi madde	0.6	Kolay madde
	Alt grup	42				
3	Üst grup	88	0.5	Çok iyi madde	0.63	Kolay madde
	Alt grup	38				
4	Üst grup	79	0.69	Çok iyi madde	0.45	Orta güçlükte
	Alt grup	10				
5	Üst grup	72	0.56	Çok iyi madde	0.44	Orta güçlükte
	Alt grup	16				
6	Üst grup	65	0.50	Çok iyi madde	0.40	Orta güçlükte
	Alt grup	15				
7	Üst grup	59	0.36	Oldukça İyi madde	0.41	Orta güçlükte
	Alt grup	23				
8	Üst grup	65	0.34	Oldukça İyi madde	0.48	Orta güçlükte
	Alt grup	31				
9	Üst grup	51	0.22**	Düzenlenip geliştirilmeli	0.40	Orta güçlükte
	Alt grup	29				
10	Üst grup	85	0.59	Çok iyi madde	0.56	Orta güçlükte
	Alt grup	26				
11	Üst grup	80	0.59	Çok iyi madde	0.51	Orta güçlükte
	Alt grup	21				
12	Üst grup	26	0.06***	Çok zayıf madde, testten çıkarılmalı	0.23	Zor madde
	Alt grup	20				
13	Üst grup	70	0.56	Çok iyi madde	0.42	Orta güçlükte
	Alt grup	14				
14	Çıkarılan 2					
15	Üst grup	59	0.40	Çok iyi madde	0.39	Zor madde
	Alt grup	19				
16	Üst grup	92	0.64	Çok iyi madde	0.60	Orta güçlükte madde
	Alt grup	28				
17	Üst grup	82	0.65	Çok iyi madde	0.50	Orta güçlükte
	Alt grup	17				
18	Üst grup	44	0.27**	Düzenlenip geliştirilmeli	0.31	Zor madde
	Alt grup	17				
19	Üst grup	86	0.70	Çok iyi madde	0.51	Orta güçlükte
	Alt grup	16				
20	Üst grup	70	0.48	Çok iyi madde	0.46	Orta güçlükte

	Alt grup	22				
21	Üst grup	96	0.72	Çok iyi madde	0.60	Orta güçlükte
	Alt grup	24				
22	Üst grup	84	0.65	Çok iyi madde	0.51	Orta güçlükte
	Alt grup	19				
23	Üst grup	42	0.20**	Düzenlenip geliştirilmeli	0.32	Zor madde
	Alt grup	22				
24	Üst grup	92	0.57	Çok iyi madde	0.63	Kolay madde
	Alt grup	35				
25	Üst grup	92	0.59	Çok iyi madde	0.62	Kolay madde
	Alt grup	33				
26	Üst grup	86	0.62	Çok iyi madde	0.55	Orta güçlükte
	Alt grup	24				
27	Üst grup	91	0.74	Çok iyi madde	0.54	Orta güçlükte
	Alt grup	17				
28	Üst grup	49	0.33	Oldukça iyi madde	0.33	Zor madde
	Alt grup	16				
29	Üst grup	78	0.54	Çok iyi madde	0.51	Orta güçlükte
	Alt grup	24				
30	Üst grup	67	0.50	Çok iyi madde	0.42	Orta güçlükte
	Alt grup	17				

** : Düzenlenip geliştirilmesi gereken madde

*** : Çok zayıf madde, testten çıkarılmalı

Tablo 8 incelendiğinde maddelerin çoğunun istenildiği gibi orta güçlükte olmakla birlikte 12. maddenin ayırt ediciliğinin çok düşük olmasından dolayı testten çıkarılması kararlaştırılmıştır.

Ön pilot uygulama sonucu elde edilen ve asıl uygulama için sonuçları kestirmeye yardımcı olan veriler Tablo 5 dikkate alınarak çok kolay ve ayırt ediciliği çok düşük olan 1. ve 14. sorular testten çıkarılmıştır. 8. ve 13. sorular ise üzerinde düzenlemeler yapılarak teste dâhil edilmişlerdir. Asıl uygulama (bkz. Tablo 8) sonucu elde edilen verilerin Tablo 5 ile karşılaştırılması sonucu 13. soruda yapılan düzenlemelerin doğru sonuçlar verdiği görülmüştür.

Ön pilot uygulama sonuçlarına göre düzenleme yapılan 9. soruda ise mevcut düzeltmelere rağmen öğrencilerin, “21 Haziran tarihinde Kuzey Yarım Kürede hiçbir yerde gölge oluşmaz” ifadesi ile ilgili sorun yaşadıkları ve bu nedenle yanlış cevaba yönelindikleri tespit edilmiştir ($r_{jx}=0.22$). Ayrıca puanlamalı bir soru olduğu için yanlış cevaba yanlış denilmesi durumunda cevaba doğru puanı verildiği için dikkat edilmesi gereken bir soru olmuştur. Verilen ifadeleri puanlamaya dayalı olan bu soru maddesi analiz sonuçlarından yola çıkarak düzeltilmeye karar verilmiştir.

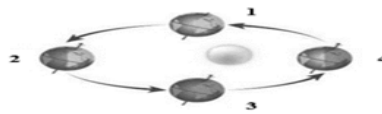
12. soruda ön pilot uygulama sonuçları analiz edildiğinde sorun tespit edilmemesine rağmen asıl uygulama sonucunda ayırt ediciliği çok düşük çıkmıştır ($r_{jx}=0.06$). Bu durum incelendiğinde pilot uygulama sonrası asıl uygulamaya hazırlık amacıyla soruya ait görsel daha anlaşılır bir görselle değiştirilmiştir. Bu sırada görsel üzerindeki ifadeler ve rakamlar düzenlenirken sehven yanlış yazıldığı için sorunun cevabı değişmiş ve cevabın seçenekler arasında bulunamamasından dolayı maddenin sağlıklı çalışmadığı görülmüştür. Normal şartlarda kullanılabilir durumda olan madde bu hata sonucu testten çıkarılmıştır.

18. sorunun ayırt ediciliği düşük çıkmış ($r_{jx}=0.27$) fakat kazanımları temsil etme özelliği bakımından tek madde olması nedeniyle geliştirilerek teste alınmasına karar verilmiştir. Tablo 9'daki veriler incelendiğinde zor bir madde olarak görülsede dikkat ölçen bir madde olduğu söylenebilir.

23. soru ile ilgili madde analizleri incelendiğinde geliştirilmesi gereken bir soru olduğu görülmüştür ($r_{jx}=0.20$). Sorunun zor olarak algılanmasından dolayı ayırt ediciliği düşük çıktığı görülmüştür. Bu durumla ilgili farklı okullardan fen bilimleri öğretmenleri ve öğrenciler ile görüşmeler yapılmış; konu anlatımı sırasında benzer sorulara değinilmemesinden kaynaklanan bir eksiklik olduğu tespit edilmiştir. Öğretim programında tek bir kazanıma bağlı olarak konuların işlenmesinin öğretmen ve öğrenciler arasında farklılıklar doğurması olağan olarak değerlendirilmektedir. Daha soyut olması gereken Dünya ve Güneş'in konumları ile ilgili sorularda öğrenciler sorunu yaşamazken, yakın çevre bilgisini ölçen bu madde de ayırt edicilik düşük çıkmıştır. Test geliştirme çalışmalarında özellikle yakın çevre ile ilgili bilgileri de ölçmek amaçlandığı için bu maddenin geliştirilerek teste kalması kararlaştırılmıştır.

Çalışmalar sonrası testin son halinde 27 maddeye yer verilmiştir. Belirtke tablosunda çıkarılan madde sonrası kazanımlarda bir değişiklik olmadığı görülmüştür.

12. Aşağıdaki görselde Dünya'nın Güneş etrafındaki hareketi gösterilmiştir



Buna göre,
I. KYK'deki bir cismin öğle saatinde gölge boyu 2 konumundan 3'e geçerken küçülür.
II. 3 konumunda Güney yarım kürede yaz mevsimi başlar.
III. 4 konumundan 1 konumuna geçerken Kuzey yarım kürede geceler kısalmır.
ifadelerinden hangileri yanlıştır?
A) I ve II. B) I ve III. C) II ve III. D) I, II ve III

Şekil 1: Çıkarılması Gereken On İkinci Madde

Şekil 1'de verilen 12. maddenin ayırt edicilik indeksi 0.06 olduğu için madde testten çıkartılmıştır.

3.3. MOBT'nin Ölçüm Güvenirliğine Ait Bulgular

MOBT'nin ölçüm güvenirliliği analizi maksadıyla hesaplanan ölçümlerin KR-20 iç tutarlılık katsayısı 0.83 olarak hesaplanmıştır. Bu değer 0.70'in üzerinde olması MOBT'nin ölçüm güvenirliliğinin iyi olduğunu ($0.7 \leq \alpha < 0.9$) kanıtlamaktadır (George ve Mallery, 2003). Bir başka ifadeyle başarı testinin geliştirilme sürecinde ölçüm güvenirliliği sağlanabilmiştir.

3.4. MOBT'nin Madde İstatistikleri

Alt ve üst grupların puan ortalamaları arasındaki farkın rastlantısal mı yoksa istatistiksel manada mı anlamlı olduğunun belirlenmesine yönelik olarak bağımsız

gruplar t-testi analizi gerçekleştirilmiştir. Elde edilen bu sonuçlar Tablo 9'da sunulmaktadır.

Tablo 9: Alt ve Üst Grupların Puan Ortalamalarına Göre Bağımsız Gruplar T-Testi Sonuçları

Soru No	Grup	N	Mean	Standart Sapma	Anlamlılık Değeri
1	Çıkarılan 1				
2	Üst grup	100	0.78	0.41	0.00
	Alt grup	100	0.42	0.36	
3	Üst grup	100	0.88	0.32	0.00
	Alt grup	100	0.38	0.48	
4	Üst grup	100	0.79	0.40	0.00
	Alt grup	100	0.10	0.31	
5	Üst grup	100	0.72	0.45	0.00
	Alt grup	100	0.16	0.36	
6	Üst grup	100	0.65	0.47	0.00
	Alt grup	100	0.15	0.35	
7	Üst grup	100	0.59	0.49	0.00
	Alt grup	100	0.23	0.42	
8	Üst grup	100	0.65	0.47	0.00
	Alt grup	100	0.31	0.46	
9	Üst grup	100	0.51	0.50	0.00
	Alt grup	100	0.29	0.45	
10	Üst grup	100	0.85	0.35	0.00
	Alt grup	100	0.26	0.44	
11	Üst grup	100	0.80	0.40	0.00
	Alt grup	100	0.21	0.40	
12	Çıkarılan 2				
13	Üst grup	100	0.70	0.46	0.00
	Alt grup	100	0.14	0.34	
14	Çıkarılan 3				
15	Üst grup	100	0.59	0.49	0.00
	Alt grup	100	0.19	0.39	
16	Üst grup	100	0.92	0.27	0.00
	Alt grup	100	0.28	0.44	
17	Üst grup	100	0.82	0.38	0.00
	Alt grup	100	0.17	0.38	
18	Üst grup	100	0.44	0.49	0.00
	Alt grup	100	0.17	0.37	
19	Üst grup	100	0.86	0.34	0.00
	Alt grup	100	0.16	0.37	
20	Üst grup	100	0.70	0.46	0.00
	Alt grup	100	0.22	0.41	
21	Üst grup	100	0.96	0.19	0.00
	Alt grup	100	0.24	0.42	
22	Üst grup	100	0.84	0.36	0.00
	Alt grup	100	0.19	0.39	
23	Üst grup	100	0.42	0.49	0.00
	Alt grup	100	0.22	0.41	
24	Üst grup	100	0.92	0.27	0.00

	Alt grup	100	0.35	0.47	
25	Üst grup	100	0.92	0.27	0.00
	Alt grup	100	0.33	0.47	
26	Üst grup	100	0.86	0.34	0.00
	Alt grup	100	0.24	0.42	
27	Üst grup	100	0.91	0.28	0.00
	Alt grup	100	0.17	0.37	
28	Üst grup	100	0.49	0.50	0.00
	Alt grup	100	0.16	0.37	
29	Üst grup	100	0.78	0.41	0.00
	Alt grup	100	0.24	0.43	
30	Üst grup	100	0.67	0.47	0.00
	Alt grup	100	0.17	0.38	

Tablo 9 incelendiğinde nihai teste ait toplam 27 sorunun her biri için üst grup ve alt grup ortalama puanları arasında anlamlı farklılıklar tespit edilmiştir. Test geliştirme çalışmaları sırasında yapılan ön pilot test ve nihai teste ilişkin güvenilirlik analizleri Tablo 10’da verilmektedir.

Tablo 10: MOBТ Güvenirlik Analizi Sonuçları

Aşama	Madde	N	Ort. Güçlük	Ort. Ayırt Edicilik	KR-20
Ön Pilot Test	30	34	0.60	0.44	0.80
Pilot Test	28	347	0.44	0.50	0.83

Tablo 10’da görüldüğü üzere MOBТ’nin güçlüğü orta düzeydedir. Bir testin güçlük indeksinin orta düzeyde olması ise test geliştirme çalışmalarında istenilen bir durumdur (Ascalon, Meyers, Davis ve Smits, 2007; Tekin, 2010). Testin ayırt ediciliği ise oldukça yüksektir. Kuder-Richardson 20 (KR-20) güvenilirlik katsayısı, madde sayısı 27’ye düşürüldüğünde de değişmemiştir.

5. SONUÇ, TARTIŞMA ve ÖNERİLER

Yapılan bu çalışma ile FBDÖP kapsamında, 8. sınıf düzeyinde mevsimlerin oluşumu konusuna yönelik geçerli ve güvenilir bir test geliştirilmeye çalışılmıştır. Bu amaçla test hazırlanırken tüm süreç boyunca geçerlik, güvenilirlik ve madde analizi çalışmaları gerçekleştirilmiştir.

Testin geçerlik çalışmaları çerçevesinde öncelikle kapsam geçerliğinin sağlanabilmesi için belirtke tablosu hazırlanmıştır (DiDonato-Barnes, Fives ve Krause, 2014; Lambie, Blount ve Mullen, 2017). Alanyazındaki mevcut testlerin sadece %8 oranında üst düzey düşünme becerilerini ölçmeye yönelik maddelerden hazırlanması genel olarak eleştirilmektedir (Gündüz, 2009). Üst düzey düşünme becerilerine sahip madde yazımının yetersizliği üzerine yapılan eleştiriler nedeniyle belirtke tablosunda üst düzey düşünme becerilerine sahip maddelere ağırlık verilmeye çalışılmıştır (bkz. Tablo 3). Bunun için Bloom’un sınıflamasıyla büyük ölçüde benzerlik göstermekle birlikte alternatif tek boyutlu sınıflamalardan farklı olarak üst düzey basamaklara ağırlık veren Haladyna’nın (1997) yapmış olduğu sınıflama kullanılmıştır. Ayrıca 2018 yılında değişen LGS sınav sistemi nedeniyle yeni nesil sorulardan oluşan etkili bir test geliştirilmeye çalışılmıştır. Bir testte

bulunan soruların sadece kazanımı kapsamı veya bilişsel alan basamaklarını örnekleme yeterli olmadığından madde türü çeşitlendirilmeye çalışılmıştır. Çünkü aynı kazanım ve aynı bilişsel alanı kapsayan farklı soru tipleri yazılabilmektedir. Örneğin bir madde hedef kazanımla beraber grafik yorumlama yeteneğini de ölçerken, başka bir madde hedef kazanımın yanında dikkati de ölçmektedir. Bazıları düz metin şeklinde kısaltılarak verilirken bazıları da görsellerle ve ön bilgilerle uzatılabilmektedir. Bu anlamda bağlam temelli sorulara da yer verilmiştir.

Testin geçerlik çalışmaları çerçevesinde yapı geçerliği de sağlanmaya çalışılmıştır. Bu bağlamda faktör analizi yapılarak tetrakorik korelasyon katsayısı hesaplanmış ve testin yapı geçerliğinin sağlandığı görülmüştür (bkz. Tablo 7). Ölçme aracında bulunan her bir maddenin testin amacına hizmet ettikleri ve kendi aralarında bir bütünlük oluşturdukları tespit edilmiştir (Büyüköztürk, 2002).

Geçerli bir ölçme aracının bir diğer önemli özelliği güvenilir olmasıdır. Güvenirlik, örneklemden örnekleme değiştiği için (Capraro ve Capraro, 2002; Henson, Kogan ve Vacha-Haase, 2001) testlerin özelliğinden çok ölçümlerin özelliğini yansıtmaktadır. Burada tespit edilmesi gereken de ölçüm güvenirligidir (Capraro, Capraro ve Henson, 2001 ve Thompson, 1994). Bundan dolayı MOBT'ni geliştirme çalışmasında ölçüm güvenirligi üzerine yoğunlaşarak örneklem için seçilen okullardaki öğretmen ve öğrenciler bilgilendirilerek gerekli bilinç oluşturulmaya çalışılmıştır. Bu sayede teste ait KR-20 güvenirlilik verileri 12. soru çıkarıldıktan sonra oldukça tatmin edici bulunmuştur ($\alpha=0.83$).

Teste madde analizleri sonucu düzeltilmesi gereken 9. 18. ve 23. sorular incelendiğinde ayırt ediciliklerinin düşük çıkma sebebi olarak sorulara ait kazanımların salgın döneminde yeteri kadar derinlemesine öğrenilememiş olduğu düşünülmektedir. Bundan dolayı kazanımlar açısından önemli olan bu sorular düzenlenerek teste dâhil edilmişlerdir.

Mevsimler konusunda oluşturulan bu başarı testi, tarama modeliyle geliştirildiği için evrene genellenebilirliği açısından, ortaokul 8. sınıf öğrencilerine uygulanabilir nitelikte bir ölçek olduğu ifade edilebilir. Yapılan çalışmada seçilen konu itibarıyla alanyazında karşılaşılmayan bir durum olması nedeniyle bu konuda önemli bir boşluğu dolduracağı düşünülmektedir. Araştırma sonucunda, ortaokul sekizinci sınıf öğrencilerinin mevsimler konusuna yönelik anlayışlarını ölçmek amacıyla 27 çoktan seçmeli sorudan oluşan, geçerliği (bkz. Tablo 7) ve ölçüm güvenirliliği sağlanmış (bkz. Tablo 9) bir test geliştirilmiştir (bkz. Ek A). Bu bağlamda, klasik test teorisinin aşamaları (Haladyna, 2004: 14, 15) dikkate alınarak hazırlanan testin, 8. sınıf öğrencilerinin mevsimler konusu ile ilgili anlayışlarını ölçmek amacıyla kullanılabilmesi söylenebilir. Bu başarı testinin geliştirilme sürecinde geçerlik ve güvenirlilik ile ilgili çalışmalardan elde edilen bulgular bağlamında aşağıda birtakım öneriler getirilmiştir:

1. Sonuçları itibariyle yüksek kapsam geçerliğine sahip olan bu test, öğrencilerin hedef kazanıma ne kadar sahip olduklarını görmesi için çok güçlü bir geri dönüt sağlayacağı gibi sorarken öğretebilme özelliğinden dolayı da öğretmenler tüm maddeleri, kesinlikle uygulamalıdır.
2. Test geliştirirken kazanıma ve taksonomiye bağlı kalmakla birlikte maddelerin farklı şekil, yapı ve özellikleri gözetilerek çeşitli becerileri ölçecek şekilde hazırlanması çalışmayı daha kapsamlı hale getirebilir.
3. Test geliştirme çalışmaları sürecinde kazanımların kapsam geçerliği göz önünde bulundurularak madde seçimi yaparken konu hakkında kavram yanılgıları üzerine yapılan araştırmaların da incelenmesi hazırlanan testin geçerliğini ve niteliğini artıracaktır.

KAYNAKÇA

- Akbulut, H. İ. ve Çepni, S. (2013). Bir üniteye yönelik başarı testi nasıl geliştirilir? İlköğretim 7. sınıf kuvvet ve hareket ünitesine yönelik bir çalışma. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 2(1), 18-44.
- Alkış, S. (2006). İlköğretim sekizinci sınıf öğrencilerinin mevsimlerin oluşumuyla ilgili fikirlerinin incelenmesi. *Marmara Coğrafya Dergisi*, (14), 107-120.
- Andrew D. P. S., Pedersen P. M. ve McEvoy C. D. (2011). *Research Methods in Sport Management*. Champaign: Human Kinetics.
- Ascalon, M. E., Meyers, L. S., Davis, B. W. ve Smits, N. (2007). Distractor similarity and item-stem structure: Effects on item difficulty. *Applied Measurement in Education*, 20(2), 153-170.
- Aslan, K., Karamustafaoğlu, O. ve Kurt, M. (2018). Otizmliler için Bir Eğitsel Oyunla “İç Organlarını Tanıyabilme” Konusunun Öğretimi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 18(4), 1887-1902.
- Atwood, R. K. ve Atwood, V. A. (1996). Preservice elementary teachers' conceptions of the causes of seasons. *Journal of Research in Science Teaching*, 33(5), 553-563.
- Balcı, A. (2021). *Sosyal bilimlerde araştırma yöntem, teknik ve ilkeler* (15. Baskı). Ankara: Pegem Akademi Yayıncılık.
- Baran, H. (2018). Açıköğretim sınavlarının faktör yapısının incelenmesi: Uluslararası İlişkiler Kuramları-I dersi örneği. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 4(1), 23-46.
- Berliner, D. C. (2020). The implications of understanding that PISA is simply another standardized achievement test. In *Handbook of Education Policy Studies* (pp. 239-258). Springer, Singapore.
- Bloom, B. S. (1956). Taxonomy of educational objectives, the classification of educational goals. New York: David McKay Company Inc. discrimination indices. *Journal of Educational Measurement*, 16(2), 109-118.
- Bolat, M., Altınbaş, A. (2018). Fen Bilgisi ve Sosyal Bilgiler Öğretmen Adaylarının Mevsimler Konusundaki Bilgi Düzeylerinin Paralel Karma Yöntem ile Belirlenmesi. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 9(16), 950-983.
- Bryman, A. ve Bell, F. (2015). *Business research methods* (6th ed.). Oxford, UK: Oxford University Press.
- Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve uygulamada eğitim yönetimi*, 32(32), 470-483.
- Capraro, M. M., Capraro, R. M. ve Henson, R. K. (2001). Measurement error of scores on the mathematics anxiety rating scale across studies. *Educational and Psychological Measurement*, 61(3), 373-386.

- Capraro, RM ve Capraro, M.M. (2002). Myers-briggs tipi gösterge puanı güvenilirliği: Bir meta-analitik güvenilirlik genelleme çalışmasını inceler. *Eğitimsel ve Psikolojik Ölçme*, 62 (4), 590-602.
- Choi, Y. ve Park, H. (2016). Shadow education and educational inequality in South Korea: Examining effect heterogeneity of shadow education on middle school seniors' achievement test scores. *Research in Social Stratification and Mobility*, 44, 22-32.
- Demirkuş, N. (2019). *Öğrenme öğretim teoriler ve yaklaşımlar* <http://www.biyoloji.egitim.yyu.edu.tr/ders/omk.html> Erişim: 27.09.2020.
- DiDonato-Barnes, N., Fives, H. ve Krause, E. S. (2014). Using a Table of Specifications to improve teacher-constructed traditional tests: an experimental design. *Assessment in Education: Principles, Policy and Practice*, 21(1), 90-108.
- Doğan, N. (2009). Ölçme araçlarını sınıflama çabaları. H. Atılğan (Ed.). *Eğitimde ölçme ve değerlendirme içinde*, 119-144. Ankara: Anı.
- Ergin, Y. D. (1995). Ölçeklerde geçerlik ve güvenilirlik. *M.Ü. Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, 7, 125-148.
- Frede, V. (2006). Pre-service elementary teacher's conceptions about astronomy, *Advances in Space Research*, 38, 2237-2246.
- George D. ve Mallery P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Boston: Allyn and Bacon.
- Gibson, R. ve Mitchell, M. (2008). *Danışmanlık ve rehberliğe giriş*. Upper Saddle Nehri: Pearson Education.
- Gönen, S., Kocakaya, S. ve Kocakaya, F. (2011). Dinamik konusunda geçerliği ve güvenilirliği sağlanmış bir başarı testi geliştirme çalışması, *Yüzüncü Yıl Üniversitesi, Eğitim Fakültesi Dergisi*, 8(1), 40-57.
- Gülen, S. (2019). Analysis of eighth grade students' academic achievements in seasons and climate unit. *Journal of Educational And Instructional Studies In The World*. 9, 1-11.
- Güler, N. (2012). *Eğitimde Ölçme ve Değerlendirme* (4. Baskı). Ankara: Pegem Akademi.
- Güler, N. (2017). *Eğitimde Ölçme ve Değerlendirme* (10. Baskı). Ankara: Pegem Akademi.
- Gündüz, Y. (2009). İlköğretim 6, 7 ve 8. Sınıf Fen ve Teknoloji Sorularının Ölçme Araçlarına ve Bloom'un Bilişsel Alan Taksonomisine Göre Analizi. *Yüzüncü Yıl Üniversitesi, Eğitim Fakültesi Dergisi*. 6(2), 150-165.
- Öksüz, Y., Demir, E. G. (2019). Açık uçlu ve çoktan seçmeli başarı testlerinin psikometrik özellikleri ve öğrenci performansı açısından karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 34(1), 259-282.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Routledge.
- Haladyna, T. M. (1997) *Writing Test Items to Evaluate Higher order Thinking*. Boston: Allyn and Bacon.
- Hasançebi, B., Terzi, Y. Küçük, Z. (2020). Madde güçlük indeksi ve madde ayırt edicilik indeksine dayalı çeldirici analizi. *Gümüşhane Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 10(1), 224-240.
- Henson, R. K., Kogan, L. R. ve Vacha, Haase, T. (2001). A Reliability Generalization Study of the Teacher Efficacy Scale and Related Instruments. *Educational and Psychological Measurement*, 61, 404-420.
- Jiun, L. T., Kamarudin, N., Talib, O. ve Hassan, A. (2018). The Effect of Structured Inquiry-Based Teaching on Biology Students' Achievement Test. *International Journal of Education, Psychology and Counseling*, 3(12), 81-89.
- Karamustafaoglu, S. ve Tutar, M. (2016). 6. Sınıf Dünya'mız, Ay ve Yaşam Kaynağımız Güneş Ünitesi'ne yönelik bir başarı testi geliştirme. *Pegem Atıf İndeksi*, 303-320.
- Karasar, N. (1991). *Bilimsel araştırma teknikleri*. Ankara: Sanem Matbaacılık.

Karslı, F.ve Ayas, A. (2013). Fen Bilgisi Öğretmen Adaylarının Kimya Konularında Sahip Oldukları Alternatif Kavramlar. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, 7(2), 284-313.

Khairani, A. Z. ve Shamsuddin, H. (2016). Assessing Item Difficulty and Discrimination Indices of Teacher-Developed Multiple-Choice Tests. In *Assessment for Learning Within and Beyond the Classroom* (pp. 417-426). Springer, Singapore.

Kızılcapan, O. ve Bektaş, O. (2018). Yedinci sınıf maddenin yapısı ve özellikleri ünitesi başarı testi geliştirilmesi: geçerlik ve güvenirlik çalışması, *The Journal of International Lingual, Social and Educational Sciences*, 4(2), 186-202.

Kikas, E. (1998). The impact of teaching on students' definitions and explanations of astronomical phenomena. *Learning and Instruction*, 8(5), 439-454.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. New York: Guilford Press.

Korkmaz, H. ve Kaptan, F. (2005). Fen eğitiminde öğrencilerin gelişimini değerlendirmek için elektronik portfolyo kullanımı üzerine bir inceleme. *The Turkish Online Journal of Educational Technology*, 4(1), 101-106.

Küçükahmet, L. (2002). *Öğretimde planlama ve değerlendirme* (13. baskı). Ankara: Nobel.

Lambie, G. W., Blount, A. J. ve Mullen, P. R. (2017). Establishing content-oriented evidence for psychological assessments. *Measurement and Evaluation in Counseling and Development*, 50(4), 210-216.

Levesque, R. J. (Ed.). (2011). *Encyclopedia of adolescence*. Springer Science and Business Media.

McMillan, J. ve Schumacher, S. (2010). *Eğitimde araştırma: Kanıta dayalı araştırma*. Upper Saddle Nehri: Pearson Education.

MEB (2018). *İlköğretim kurumları (ilkokullar ve ortaokullar) fen bilimleri dersi (3, 4, 5, 6, 7 ve 8. sınıflar) öğretim programı*. Ankara: MEB Yayınevi.

Merenda, P. F. (2007). Update on the decline in the education and training in psychological measurement and assessment. *Psychological reports*, 101(1), 153-155.

Newfields, T. (2006, May). Teacher development and assessment literacy. *At presented Proceedings of the 5th Annual JALT Pan-SIG Conference. Shizuoka, University College of Marine Science, Tokai.*

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

Pandra, V. ve Mardapi, D. (2017). Development of Mathematics Achievement Test for Third Grade Students at Elementary School in Indonesia. *International Electronic Journal of Mathematics Education*, 12(3), 769-776.

Rollins, M. M., Denton J., J. ve Janke D. L. (1983). Attainment of Selected Earth Science Concepts by Texas High School Seniors. *Journal of Education-al Research* 7, 81- 88.

Sadler, P. M. (1992). *The initial knowledge state of high school astronomy students*, Harvard University, Cambridge, MA. PhDT, 53(5A), 1470.

Schoon, K. J. (1992). Students' alternative conceptions of earth and space. *Journal of Geological Education*, 40(3), 209-214.

Sharp, J. G. (1996). Children's astronomical beliefs: a preliminary study of Year 6 children in south-west England. *International. Journal of Science Education*, 18(6), 685-712.

Sneider, C., Bar, V. ve Cavanagh, C. (2011). Learning about seasons: A guide for teachers and curriculum developers. *Astronomy Education Review*, 10(1).

- Sontay, G. ve Karamustafaoğlu, O. (2020). Fen bilimleri dersi “Güneş, Dünya ve Ay” ünitesine yönelik başarı testinin geliştirilmesi. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 40(2), 511-551.
- Struyven, K., Dochy, F. ve Janssens, S. (2005). Students’ perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education*, 30(4), 325-341.
- Tavşancıl, E. (2010). *Tutumların ölçülmesi ve SPSS ile veri analizi*. Ankara: Nobel Yayın Dağıtım
- Tekin, H. (1991). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Yargı.
- Tekindal, S. (Ed). (2017). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Akademi.
- Tezbaşaran, A. A. (2008). *Likert tipi ölçek hazırlama kılavuzu*. Ankara: Türk Psikologlar Derneği.
- Tiruchittampalam, S., Nicholson, T., Levin, J. R. ve Ferron, J. M. (2018). The effects of preliteracy knowledge, schooling, and summer vacation on literacy acquisition. *The Journal of Educational Research*, 111(1), 28-42.
- Thompson, B. (1994). Guidelines for Authors. *Educational and Psychological Measurement*, 54, 834-47.
- Trumper, R. (2001a). A Cross-age study of senior high school students’ con-ceptions of basic astronomy concepts, *Research in Science and Technological Education*, 19, 1.
- Trumper, R. (2001b). A cross-age study of senior high school students’ conceptions of basic astronomy concepts. *Research in Science and Techno-logical Education*, 19(1), 97-109.
- Turgut, M. F. ve Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme*. Pegem Akademi.
- Türk, C., Alemdar, M. ve Kalkan, H. (2012). Mevsimleri anlama seviyesinin belirlenmesi ilköğretim öğrencilerinin konusu. *Dünyada Eğitim ve Öğretim Araştırmaları Dergisi*, 2 (1), 62-67.
- Türk, C., Kalkan, H., Kiroğlu, K. ve Ocak İskeleli, N. (2016). Elementary school students' mental models about formation of seasons: A cross sectional study. *Journal of Education and Learning*, 5(1), 7-30.
- Türk, C. ve Kalkan, H. (2015). The effect of planetariums on teaching specific astronomy concepts. *Journal of Science Education and Technology*, 24(1), 1-15.
- Uygur, E. (2009). *İlköğretim 7. Sınıf Fen ve Teknoloji Dersi Kuvvet ve Hareket Ünitesinin Öğretiminde İşbirlikçi Öğrenme Yönteminin Öğrenci Başarısına, Tutuma ve Bilgi Kalıcılığına Etkisi*. Yayınlanmamış Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Yancı, M.V. (2020). *Ortaokul ve İmam Hatip Ortaokulu Fen Bilimleri 8. Sınıf Ders Kitabı*. Ankara: Dikey Yayıncılık.
- Yazıcıoğlu, S. (2017). *Oyun temelli etkinliklerin ortaokul 6. sınıf öğrencilerinin akademik başarılarına, fen öğrenmeye yönelik motivasyonlarına ve tutumlarına etkisi: Işık ve ses ünitesi örneği*. (Yayınlanmamış Yüksek Lisans Tezi), Giresun Üniversitesi, Giresun.
- Yıldırım, C. (1999). *Eğitimde Ölçme ve Değerlendirme*. Ankara: ÖSYM.
- Yin, R. K. (2009). *Case Study Research: Design and Methods* (4th ed). Thousand Oaks,CA: Sage.

EKLER

Ek 1. Mevsimlerin Oluşumu Başarı Testi Erişim Bilgileri

Kısa URL

QR KOD

<https://qrco.de/bd1y5j>

