



Başarı Testi Geliştirme Süreci: İlkokul Dördüncü Sınıf Maddeyi Tanıyalım Ünitesi Örneği

The Phases of Achievement Test Development: The Case of Fourth-Grade Introduction to Matter Unit

Gökşen ÜÇÜNCÜ¹, Gönül SAKIZ²

Öz

Başarı testleri öğrencilerin akademik performanslarını ölçmek için en yaygın olarak kullanılan veri toplama araçları arasında yer alırlar. Bu nedenle, seçili disiplinlerde öğrenci kazanımlarını kusursuz şekilde ortaya koyabilen geçerli ve güvenilir başarı testlerinin varlığı büyük önem arz eder. Fakat, başarı testi geliştirme üzerine yapılan araştırmalar oldukça kısıtlıdır. Bu araştırmanın amacı, bilimsel araştırmalarda kullanılacak bir başarı testinin geliştirilmesi sürecinde izlenmesi gereken adımları örnek bir fen bilimleri ünitesi (Maddeyi Tanıyalım) çerçevesinde sergilemek ve geliştirilen başarı testini literatüre kazandırmaktır. Araştırmanın ilk bölümünde, seçili üniteye ilişkin 11 kazanım için 35 çoktan seçmeli soru maddesi hazırlanmıştır. Webb'in (1997) öğrenme kazanımlarının değerlendirilmesine ilişkin belirlediği ölçütlere dayanarak alan uzmanları ve deneyimli öğretmenlerden alınan görüşler Lawshe tekniği (1975) ile değerlendirilmiştir. Çalışmanın ikinci bölümünde ise oluşturulan ölçek formu İstanbul'un Beykoz İlçesi'nde yer alan iki ilkokulda dört dördüncü sınıf şubesinde öğrenim gören 84 öğrenciye uygulanmıştır. ITEMAN programı ile yapılan madde analizi testin geçerli olduğunu göstermiştir. Pearson korrelasyon ve Spearman Brown güvenilirlik katsayıları sırasıyla .85 ve .83 olarak bulunarak testin güvenilir olduğunu ortaya koymuştur. Geliştirilen testin, fen eğitimi alanına katkı sağlaması ve yapılacak ilişkili araştırmalarda kullanılması beklenmektedir. Açıklanan prosedürlerin çeşitli branşlarda geliştirilecek geçerli ve güvenilir başarı testlerinin oluşturulma sürecine rehberlik edeceğine inanılmaktadır.

Anahtar Kelimeler: başarı testi geliştirme, Lawshe tekniği, ITEMAN analizi, fen bilimleri dersi.

Abstract

Achievement tests are major data collection tools used to measure students' academic performance in any given field. Therefore, valid and reliable achievement tests reflecting the precise measurement of expected outcomes in selected disciplines are highly needed. However, research on achievement test development is limited. The purpose of this research was to present the phases of an achievement test development using the sample case of a science lesson unit, Introduction to Matter. Initially, in the development process, based on 11 objectives of the unit, 35 multiple-choice questions were prepared. Grounded on Webb's (1997) criteria for the evaluation of the learning objectives, the experts' and teachers' opinions were collected and the results were assessed using Lawshe technique (1975). In the second part, the scale was applied to 84 fourth-grade students in four classrooms of two elementary schools located in Beykoz, Istanbul. The item analysis conducted using ITEMAN program showed that the scale was valid. Item difficulties were calculated. Pearson correlation coefficient and Spearman-Brown reliability coefficient were found .85 and .83, respectively, confirming that the test was reliable. It is expected that the developed test would contribute to the field of science education and be used in related research. The procedures explained in the paper hopefully provide guidance for researchers for the development of more valid and reliable achievement tests in various disciplines.

Keywords: achievement test development, Lawshe technique, ITEMAN analysis, science lesson.

1 M.E.B. Ölçme, Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü, Ankara, Türkiye, <https://orcid.org/0000-0001-8107-229X>

2Marmara Üniversitesi Temel Eğitim- Sınıf Öğretmenliği ABD, Kastamonu, Türkiye, <http://orcid.org/0000-0002-7095-9554>

Atıf / Citation: Üçüncü, G. ve Sakız, G. (2020). Başarı Testi Geliştirme Süreci: İlkokul Dördüncü Sınıf Maddeyi Tanıyalım Ünitesi Örneği. *Kastamonu Education Journal*, 28(1), 82-94. doi:10.24106/kefdergi.3440

Extended Abstract

Purpose and Significance: The purpose of this research was to present the phases of an achievement test development using the sample case of Introduction to Matter Unit in fourth grade science lessons. Achievement tests are major data collection tools used to measure students' academic performance in any given field. Therefore, valid and reliable achievement tests reflecting the precise measurement of expected outcomes in selected disciplines are highly needed. This research will hopefully encourage the development of more valid and reliable achievement tests in various fields.

Methods and Procedure: Initially, the standards of fourth grade Introduction to Matter unit were examined and 35 multiple-choice questions were prepared. Then, the experts and the experienced teachers examined the questions based on the science curriculum and Webb's criteria (1997). The results were evaluated using Lawshe technique (1975) and 35 items were reduced to 32 items. These questions were applied to 84 of fourth grade students and ITEMAN program was used for analyzing the items.

Results: Two items were found invalid. These items were removed from the test. In the final phase, 30 items remained in the test. The item analysis conducted using ITEMAN program showed that the scale was valid. The item difficulties were calculated. Pearson correlation coefficient and Spearman-Brown reliability coefficient were found .85 and .83, respectively, confirming that the test was reliable.

Discussion and Conclusions: In this study, a reliable and valid achievement test consisted of 30 multiple-choice questions was developed. The developed achievement test would hopefully contribute to the field and be used in the related research. The procedures explained in this paper are expected to provide guidance for the researchers to develop more precise, valid and reliable achievement tests in different fields.

1. Giriş

Ölçme ve değerlendirme eğitim programlarının temel öğelerindedir. Ölçme, herhangi bir gözlem sonucunu sayı veya sıfatlarla ifade etme, değerlendirme ise ölçme sonuçlarını belli bir ölçüte vurarak ölçülen nitelik hakkında bir yargıya ulaşma çabasıdır (Turgut ve Baykul, 2014). Ölçme ve değerlendirme, eğitim sisteminde bireylere hangi girdilerin ne ölçüde kazandırıldığını ya da kazandırılmadığını ortaya koyar. Bu anlamda günümüzde öğrencilerin başarılarının değerlendirilmesi ve gerekli yönlendirmelerin yapılabilmesi bu iki ölçüte bağlıdır. Okullarda kullanılan testlerin genel olarak iki amacı vardır. Bu amaçlardan biri öğrencilerin belirli bir üniteye veya konudaki eksikliklerinin belirlenmesidir. Buna “öğrenmelerin izlenmesi” denilmektedir. Diğer amaç ise öğrencilerin belirli bir ünite, konu ya da üniteler topluluğundaki düzeylerini ve başarılarını ölçmektir. Bu işlem ise “erişiyi belirleme” olarak adlandırılır (Özçelik, 2010).

İzleme ya da erişiyi belirlemek için geliştirilen testlerde ölçülecek davranışları belirlemek oldukça önemlidir. Öğretim programlarının etkililiği ve öğrencilerin başarılarını doğru şekilde ölçmek ve yorumlamak öncelikle kazanımlara uygun soruların hazırlanması ile mümkündür. Ancak, her kazanım ölçülecek bir davranış ifade etmeyebilir. Bunun için, kazanımlar incelenip, davranışları gözlemlenebilir göstergeler olarak ifade etmek gerekmektedir. Can Şen ve Eryılmaz (2011), lise öğrencileri için basit elektrik devreleri konusuna yönelik başarı testi geliştirme çalışmalarında kazanımların incelenip, her bir kazanımın sorularla açık olarak ölçülebilmesinin gerekli olduğunun üzerinde durmaktadır. Alanyazında, öğretim hedeflerinin sınıflanmasına yönelik olsa da ölçmeye ağırlık vermesi nedeniyle, soru hazırlamada kazanımlardaki bilişsel düzeylerin belirlenmesi için yaygın olarak Bloom taksonomisinden faydalanılmaktadır (Dindar ve Demir, 2006). Bloom taksonomisi, zihinsel gelişim düzeyini dikkate alarak basitten karmaşığa doğru altı aşamadan oluşmaktadır. Bu aşamalar, hatırlama/bilgi, anlama/kavrama, uygulama, analiz, değerlendirme ve yeniden oluşturma/sentez olarak tanımlanmıştır. Hatırlama/bilgi aşamasında öğrencinin bir kavramı ya da olguyu anlamını bilerek söylemesi veya ezberden söylemesi vardır. Anlama/kavrama aşamasında bilgi düzeyinde kazanılan davranışın öğrenci tarafından örneklendirilmesi, benzer bir durumla bu bilgiyi ilişkilendirmesi, bilinen iki durumu karşılaştırması beklenir. Uygulama aşamasında, elde edilen bilgilerin kullanılarak bir uygulama yapılması, bir problemin çözülmesi beklenir. Analiz düzeyinde, bir bilgi bütünü o bütünü içinde yer alan biçimiyle öğelerine ayırma ve bu öğeler arası ilişkileri görme işidir. Belirli kriterlere göre bir yargıya varma süreci değerlendirme, parçaları bir araya getirerek özgün bir eser oluşturma, üretme ise yeniden oluşturma sürecidir (Ayyacı ve Türkdoğan, 2009).

Başarı testlerinin geliştirilmesinde test ile yoklanacak kazanımlar ve bu kazanımların ifade ettiği bilişsel düzeylerine uygun soru tiplerini de belirlemek gerekmektedir. Öğrencilerin bir konu üzerinde öğrenmesi beklenen davranışları ölçmek için oluşturulan testler, anket, açık uçlu sorular, doğru-yanlış tipi, kısa cevaplı, boşluk doldurmalı ve çoktan seçmeli sorular gibi farklı türlerde hazırlanabilir (Akbulut ve Çepni, 2013; Balcı, 2011; Turgut ve Baykul, 2014). Uygulama ve değerlendirme kolaylığı sağlayan çoktan seçmeli sorular, bu kolaylığa sahip doğru-yanlış, boşluk doldurma ve kısa cevaplı soru türlerine göre bilişsel alan basamaklarından bilgi ve kavramanın üzerine çıkabilmektedir (Özçelik, 2010). Ayrıca, çoktan seçmeli başarı testlerinin içerdiği soru sayısının fazla olması verilen kazanımların tamamını kısa sürede yoklama faydası sağlamaktadır. Çoktan seçmeli soru tipindeki başarı testleri tüm bilgileri ölçme ve değerlendirme imkanı sunduğu için diğer ölçme ve değerlendirme araçlarına göre en fazla tercih edilen ölçme aracıdır (Akbulut ve Çepni, 2013). Yurt içi ve yurtdışı eğitim-öğretim sistemlerinde öğrencilerin bilişsel seviyelerini ölçmek ve değerlendirmek için çoktan seçmeli testler sıklıkla kullanılmaktadır (Pressley, Yokoi, Van Meter, Van Etten ve Freebern, 1997). Pek çok eğitim kurumu, bir üst eğitim kurumuna öğrenci seçme ve yerleştirme için öğretim programlarının kazanımları ile uyumlu çoktan seçmeli testlerden faydalanmaktadır (Brown ve Conley, 2007; Edwards, 2010; Kara ve Çepni, 2011; Yan ve Erduran, 2008).

Alan yazında akademik başarıyı ölçmeye yönelik geliştirilen başarı testlerinde genellikle çoktan seçmeli sorular tercih edilmiştir (Açıkgöz ve Karslı, 2015; Akbulut ve Çepni, 2013; Akdeniz, Öztürk ve Bakırcı, 2017; Bakırcı, Artun ve Şenel, 2016; Bozdoğan ve Kavcı, 2016; Can Şen ve Eryılmaz, 2011; Can ve Yıldırım, 2017; Çalık ve Ayas, 2003; Gönen, Kocakaya ve Kocakaya, 2011; Korkmaz ve Konukaldı, 2015; Şahin ve Ercan, 2015). Ancak, Kutlu, Yalçın ve Pehlivan (2010) dördüncü sınıf maddeyi tanıyalım ünitesi ile ilgili yapmış oldukları kazanıma dayalı soru yazma ve puanlama çalışmasında, kazanımların odaklandığı becerileri göz önünde bulundurarak, bu becerilerin ölçülmesinde açık uçlu soruları tercih etmişlerdir. Kazanımların zihinsel beceri boyutunda yorumlanması ve özellikle problem çözme, eleştirel düşünme ve yaratıcı düşünme gibi becerilerin yoklanmasının hedeflendiği durumlarda, açık uçlu soruların bu amaca çoktan seçmeli sorulara göre daha iyi hizmet ettiği söylenebilir. Bu nedenle çoktan seçmeli soruları tercih ederken beceri boyutu da göz ardı edilmemelidir. Buradan çıkarımla, testin uzunluğu da göz önüne alınarak, önemli görülen

davranışlar belirlenip, bu davranışların bilişsel düzeyine uygun ve her davranışa yönelik en az bir soru hazırlanmalıdır (Balcı, 2011; Özçelik, 2010).

Bir başarı testinin geliştirilmesinde kazanımlar ile bu kazanımların bilişsel düzeylerinin belirlenmesi ve soru tipinin tercih edilmesinden sonra belirli aşamalar izlenmesi gerekmektedir. Ayhan (2010) başarı testi geliştirme sürecinde, test kapsamına alınacak davranışların belirlenmesi, bu davranışlara uygun soruların hazırlanması, ön deneme test formunun oluşturulması, maddelerin geçerlik-güvenirlik çalışmasının yapılması, analiz sonuçlarına göre ana teste alınacak soruların belirlenmesi şeklinde aşamalar belirlemiştir. Alan yazına bakıldığında araştırmacıların bir üniteye ilişkin başarı testi geliştirme sürecine odaklı çalışmaları olduğu gibi, bir araştırma içerisinde yapılan uygulamanın akademik başarı üzerine etkisini incelemek adına araştırmacılar tarafından geliştirilen başarı testleri de olduğu görülmektedir. Sadece başarı testi geliştirmeye odaklı çalışmalardan biri olan Çalık ve Ayaş'ın (2003), çözeltiler konusu merkezinde geliştirdikleri kavram başarı testinde izlenen süreçler 18 adımda açıklanmıştır. Bu aşamalarda öncelikle testin içeriğini oluşturan konu ile ilgili yapılan çalışmaların bulguları paylaşılmış, daha sonra konuya ilişkin kazanım incelemeleri ve kazanıma dayalı hazırlanan madde sayısı verilmiştir. 25 soruluk hazırlanan ön form üzerinde kapsam geçerliği için yapılan çalışmalar ayrıntılı bir şekilde irdelenmiş, ön formun uygulanması ve bu uygulama sonucunda yapılan madde analizleri verilmiştir. Benzer şekilde, Can Şen ve Eryılmaz (2011), basit elektrik devreleri için geliştirdikleri başarı testi çalışmasında, kazanım listesinin oluşturulması, belirtke tablosunun hazırlanması, maddelerin oluşturulması, maddelerin alan uzmanlarınca incelenerek ön formun oluşturulması, ön formun uygulanarak madde analizlerinin yapılması ve nihai formun oluşturulması olmak üzere altı aşamalı bir yol izlemiştir. Bu çalışmada ön form oluşturulurken, testin bir anlamda kapsam ve yapı geçerliğine hizmet etmesi beklenen uzman incelemesinde, uzmanların özellikleri belirtilerek yapılanlar detaylı biçimde anlatılmıştır. Gönen, Kocakaya ve Kocakaya (2011), dinamik konusunda lise öğrencilerine yönelik geliştirdikleri başarı testi çalışmasında, geçerlik çalışması için aşamalı bir yol sunmuş olmakla birlikte, kapsam geçerliği için maddelerin iki alan uzmanı tarafından incelendiği belirtilmiştir. Akbulut ve Çepni (2013), ortaokul yedinci sınıflar kuvvet ve hareket ünitesi için geliştirdikleri başarı testinde alan yazınında yaygın olarak takip edilen aşamalara ek olarak, Webb' in (1997) öğretim programları kazanımları ile yapılan sınavların uyumunu incelemek için geliştirdiği uyum ölçütlerine dayalı bir kapsam geçerlik çalışması yürütmüştür. Bu çalışmanın standart ölçütlere dayalı olarak yapılmasının sorular ve öğretim programı kazanımları arasındaki uyumun daha yüksek olmasını sağlayacağı düşünülmektedir (Akbulut ve Çepni, 2013). Açıkgöz ve Karlı (2015) tarafından alternatif ölçme-değerlendirme teknikleri kullanılarak yedinci sınıflar iş ve enerji konusunda geliştirilen başarı testinin kapsam geçerlik çalışmasında uzman görüşünün alındığı belirtilmiş, fakat bu görüşlerle ilgili ayrıntı verilmemiştir. Bununla birlikte, Bozdoğan ve Kavcı (2016), altıncı sınıf öğrencilerine yönelik kuvvet ve hareket ünitesine ilişkin geliştirdikleri başarı testi için alan yazınında yaygın olarak tercih edilen aşamaları izlemiş; ancak testin kapsam geçerliği için yapılan çalışma uzman görüşlerinin alınması ve bu görüşlerin uyumunun belirlenmesi hususunda kısıtlı kalmıştır. Bu sınırlılık, Şahin ve Ercan (2015) tarafından geliştirilen yedinci sınıf öğrencilerine yönelik kuvvet ve hareket ünitesi başarı testinde; Korkmaz ve Konukaldı (2015) tarafından geliştirilen yedinci sınıf öğrencilerine yönelik vücudumuzda sistemler ünitesi başarı testinde; Bakırcı, Artun ve Şenel (2016) tarafından geliştirilen yedinci sınıf öğrencilerine yönelik gök cisimlerini tanıyalım başarı testinde; Akdeniz, Öztürk ve Bakırcı (2017) tarafından geliştirilen sekizinci sınıf öğrencilerine yönelik kuvvet ve hareket ünitesi başarı testinde; Can ve Yıldırım (2017) tarafından beşinci sınıf öğrencilerine yönelik maddenin değişimi ünitesi için geliştirilen başarı testinde de görülmektedir. Diğer bir ifadeyle, geliştirilen başarı testlerindeki maddelerin geçerlik çalışmalarında yaygın olarak uzman incelemesi yapılmakla birlikte uzmanların maddeler ile ilgili vermiş oldukları görüşler arasındaki uyum ve uyum yüzdesinin geçerli olup olmadığı yönünde bilgiye rastlanmamaktadır.

Bu çalışmanın başarı testi geliştirmek için belirlediği maddeyi tanıyalım ünitesi ile ilgili olarak yapılan çalışmaların çoğunlukla ortaokul düzeyinde gerçekleştirildiği görülmektedir (Ateş, 2004; Bilgin, Ay ve Coşkun, 2013; Bozkurt, 2010; Saraç, 2018) ilkökul düzeyinde dördüncü sınıflar Maddeyi Tanıyalım ünitesi için Uyanık (2014) tarafından 28 maddelik çoktan seçmeli bir test geliştirilmiştir. Bu test incelendiğinde soruların genel olarak bilgi düzeyinde kaldığı, kazanımların ifade ettiği kavrama ve uygulama becerileri yönünden yeterli olmadığı belirlenmiştir. Bu nedenle, bu çalışmada ilkökul dördüncü sınıf öğrencilerinin Fen Bilimleri dersi Maddeyi Tanıyalım Ünitesi kazanımları ile ilgili başarı düzeylerini ölçmek için geçerliği ve güvenilirliği sağlanmış çoktan seçmeli bir test geliştirilerek, test geliştirme aşamalarının tanıtılması ve geliştirilen testin literatüre kazandırılması amaçlanmıştır. Bu testin, öğrencilerin hazır bulunuşluklarını, öğrenme durumlarını belirleme ve beraberinde öğrenme etkinliklerini düzenleme konusunda öğretmenlere ve araştırmacılara yardımcı olacağı düşünülmektedir.

2. Yöntem

Örneklem

Araştırmanın örneklemini 2014-2015 yılında İstanbul İli Beykoz İlçesinde öğrenim gören 84 ilkokul dördüncü sınıf öğrencisi oluşturmaktadır. Örneklem grubu araştırmacıların kolaylıkla ulaşabileceği uygun örnekleme yöntemine göre belirlenmiştir. Örneklem grubunu oluşturan öğrencilerin yaşları ortalaması 9,46 olarak hesaplanmıştır. 84 öğrencinin 38'ini kız, 46'sını erkek öğrenciler oluşturmaktadır. Türkiye İstatistik Kurumu (2013) tarafından açıklanan 2013 yılı nüfus dağılımlarına ilişkin istatistiklere göre Beykoz İlçesi'nde 9-10 yaş aralığında yer alan bireylerde erkek nüfusunun kız nüfusundan fazla olduğu görülmektedir. Bu durum çalışmadaki katılımcıların cinsiyet dağılımı ile benzerlik göstermektedir.

Test Geliştirme Süreci

Test geliştirme sürecinin başında öncelikle testin geliştirileceği üniteye karar verilmiştir. Seçilen ünite fen bilimleri dersi Maddeyi Tanıyalım Ünitesi'dir. Bu ünite, fen bilimleri dersinde yer alan pek çok konunun temelini oluşturan ve fen bilimleri öğretmenlerinin en çok önem verdiği konuların başında gelen madde kavramına yöneliktir (Güler Demirci, 2008). Maddeyi Tanıyalım Ünitesi, içerdiği kavramlar ve uzunluğu nedeni ile öğrencilerin anlamada güçlük çektikleri bir ünite olarak değerlendirilmektedir.

Ölçme değerlendirme sürecinde konu kadar tercih edilen test türü de oldukça önemlidir. Öğrencilerin, madde kavramı ile ilgili edindikleri bilginin düzeyini belirlemek, öğrenme etkinliklerini düzenlemek için geçerli ve güvenilir bir başarı testi oluşturmak gereklidir. Bu açıdan bakıldığında, konuya ilişkin kazanımların ölçülmesi ancak geçerliği ve güvenilirliği yüksek bir test ile gerçekleştirilebilir. Başarıyı ölçmeye dönük yapılan çalışmalarda geçerli ve güvenilir testler oluşturulabilmesi için, kazanımların incelenmesi, her kazanıma uygun en az üç soru oluşturulması, uzman görüşü alınarak maddelerin düzenlenmesi, düzenlenmiş maddeler ile ön çalışma yapılması ve madde analizi ile teste son halini verme yolları izlenir (Akbulut ve Çepni, 2013).

Test maddeleri oluşturulurken, program kazanımları ve sorular arasındaki uyumu sağlamak için Webb'in (1997) geliştirdiği ölçütlerden üçü göz önünde bulundurulmuştur. Bunlar;

1. Kategorik birlik ölçütü (categorical concurrence): Ölçme aracındaki yer alan soruların konuya ilişkin tüm kazanımları kapsayıp kapsamadığı uzmanlar tarafından incelenir ve uyum yüzdesine bakılır.
2. Bilgi uygunluk aralığı ölçütü (depth of knowledge consistency): Kazanımlar ve soruların uyumu incelenir.
3. Temsilin dengesi ölçütü (balance of representation): Soruların kazanımlara dağılımını içeren bir ölçüttür. Bu dağılımın uygun olup olmadığını belirlemek için Denge İndeksi= $1 - \left[\frac{\sum_{k=2}^O |1/ (O) - I (K)/ H|}{2} \right]$ formülü uygulanır. Formülde O = Toplam kazanım sayısı, I (K) = Kazanım başına düşen soru sayısı, H = Toplam soru sayısını ifade etmektedir. Formülden elde edilen denge indeksi katsayısı .70'den büyük olursa, kazanımların sorulara dağılımının uygun olduğu sonucuna varılır.

Bu çalışmada Webb'in (1997) belirtilen ölçütleri esas alınarak, başarı testi geliştirme süreci, soruların geliştirilmesi, uzman görüşüne sunulması, uzman görüşleri neticesinde düzenlenen testin yapı geçerliğinin sağlanması ve sorularının kazanımlara dağılımının uygunluğunun test edilmesi olmak üzere başlıca dört aşamada gerçekleştirilmiştir.

Birinci aşama: Bu aşamada başarı testinin maddelerinin hazırlanması için öncelikle üniteye ilişkin kazanımlar incelenmiştir. Kazanımlar, taksonomik olarak, hatırlama, kavrama, uygulama, analiz, değerlendirme ve yaratma basamaklarına göre irdelenmiştir. Deneme formu için hazırlanan soruların ana test için düşünülen soru sayısının üç katı kadar maddeden oluşması yaygın olarak belirlenen bir kuraldır (Turgut ve Baykul, 2014). Bu nedenle ölçülecek her davranışın üç katı kadar soru hazırlanması anlayışıyla sekiz ana başlıkta toplanan 11 kazanımdan her birine en az üç madde hazırlanarak 35 maddelik bir test formu oluşturulmuştur. Hazırlanan soruların, kazanım ifadelerini tam olarak kapsayıp kapsamadığını test edebilmek için bir belirtke tablosu oluşturulmuştur. Bu belirtke tablosu aynı zamanda

kapsam geçerliği için de bir kanıt oluşturmaktadır (Can Şen ve Eryılmaz, 2011). Test belirtke tablosuna EK-I' de yer verilmiştir.

İkinci aşama: Bu aşamada belirtke tablosu ve hazırlanan test maddeleri bir uzman grup tarafından incelenmiştir. Uzman grubun 5 ile 40 kişi arasında değişebileceği, fakat en az beş kişi olması öngörülmektedir (Yurdugül, 2005). Bu araştırmada uzman grup bir ölçme ve değerlendirme uzmanı, üç fen bilimleri, dört sınıf öğretmeni olmak üzere sekiz kişiden oluşmuştur. Uzman grubun demografik özellikleri Tablo 1'de verilmiştir.

Tablo 1. Uzmanların Demografik Özellikleri

No	Cinsiyet	Yaş	Unvan	Branş
1	E	32	Dr.	Ölçme ve Değerlendirme
2	K	32	Dr.	Fen Öğretmeni
3	K	33	Dr.	Sınıf Öğretmeni
4	K	33	Dr.	Sınıf Öğretmeni
5	E	33	Dr.	Fen Öğretmeni
6	K	42	Öğretmen	Fen Öğretmeni
7	K	42	Öğretmen	Sınıf Öğretmeni
8	E	35	Öğretmen	Sınıf Öğretmeni

Not: E: Erkek, K: Kadın

İnceleme Webb'in (1997) ortaya koyduğu ve çalışmanın ele aldığı ilk iki ölçüte göre yapılmıştır. Böylece, uzmanlar soruların tüm kazanımları kapsayıp kapsamadığını ve soruların ilgili kazanımla uyumlu olup olmadığını değerlendirmişlerdir. Belirlenen uzman grubuna oluşturulan test maddeleri ve ünite kazanımlarını içeren listeler verilmiş, soruların belirlenen iki ölçüte göre –“tüm kazanımları kapsıyor mu?” ve “kazanımlar ile sorular uyumlu mu?”– değerlendirmeleri istenmiştir. Bu değerlendirme sonuçlarını nicel olarak ortaya koymak için Lawshe (1975) tekniğinden faydalanarak, her bir madde için kapsam geçerlik oranları (KGO) elde edilmiştir.

Lawshe tekniği (1975), hazırlanan test maddelerinin kapsam geçerliği için başvuru uzman görüşlerinin alınması çalışmasını nitel boyuttan nicel boyuta taşıyan bir yöntemdir. Bu teknikte, test maddelerinin, "hedeflenen davranışları ölçmede yeterlidir", "hedeflenen davranışları ölçmesi için düzenlenmelidir" ve "hedeflenen davranışları ölçmede yetersizdir" gibi ölçütler ile uzmanlarca değerlendirilmesi beklenir.

Lawshe (1975) tekniği altı aşamada gerçekleştirilir. Bunlar, (1) alan uzman grubu oluşturulur; (2) ölçek form oluşturulur; (3) uzman görüşleri alınır; (4) maddelere ilişkin kapsam geçerlik oranları hesaplanır; (5) kapsam geçerlik indeksi belirlenir ve (6) kapsam geçerlik oranları indeks ölçütlerine göre değerlendirilerek ölçekte yer alacak maddeler belirlenir. Uzman grubundan alınan görüşler ile aşağıda verilen formül uygulanarak kapsam geçerlik oranları belirlenir.

$$KGO = \frac{N_G}{N/2} - 1$$

N_G = “Madde gereklidir/uygundur” cevabı veren uzmanların sayısı

N = Görüş bildiren toplam uzman sayısı

Formül uygulandığında matematiksel olarak uzmanların tamamı uygundur derse $KGO = 1$, uzmanların yarısı uygundur derse $KGO = 0$, uzmanların yarısından fazlası uygundur derse $KGO > 0$ ve uzmanların yarısından azı uygundur derse $KGO < 0$ değerleri ortaya çıkacaktır. Buna göre, elde edilen bulguların değerlendirilmesinde bakılacak KGO değerleri (Veneziano ve Hoper, 1997) Tablo 2'de verilmiştir.

Tablo 2. Kapsam Geçerlilik Oranları İçin Minimum Değerler

Uzman Sayısı	Minimum KGO Değeri	Uzman Sayısı	Minimum KGO Değeri
5	0.99	13	0.54
6	0.99	14	0.51
7	0.99	15	0.49
8	0.78	20	0.42
9	0.75	25	0.37
10	0.62	30	0.33
11	0.59	35	0.31
12	0.56	40	0.29

KGO = Kapsam Geçerlilik Oranı

Belirlenen uzman grubun sekiz kişilik olduğu düşünülürse, formül sonucu elde edilen bulgular .05 anlamlılık düzeyinde .78 değeri ölçüt alınarak değerlendirilmiştir. Tek boyutlu bir testte kapsam geçerlik indeksi (KGI), deneme formuna alınacak maddelerin toplam kapsam geçerlik oranları (KGO) ortalamaları üzerinden elde edilir.

Üçüncü aşama: Bu aşamada uzman görüşleri doğrultusunda düzeltilen testin yapı geçerliğinin sağlanması için madde analizi yapılır (Baykul, 2015). Böylece soruların ve testin daha geçerli ve nitelikli olması sağlanır. Bu çalışmada, hazırlanan 35 soru belirlenen ölçütler ve uzman görüşleri doğrultusunda, 32 soruya düşürülmüştür. Daha sonra hazırlanan ön form İstanbul ili Beykoz ilçesi ilkököl dördüncü sınıfta öğrenim görmekte olan 84 öğrenciye 40 dakikalık bir ders süresince uygulanmıştır.

Elde edilen veriler doğrultusunda madde analizi yapılmıştır. Maddelerin ayırt edicilik indeksi -1 ile +1 arasında değerler almaktadır. Ayırt ediciliği yüksek maddeler testin geçerliğini artırmaktadır. Bununla ilgili olarak .40 ve üzerinde olan maddeler ayırt ediciliği çok yüksek; .30 - .39 arası oldukça iyi; .20 - .29 arası düzenlenmeli; .19 ve daha küçük değerler için ise testten çıkarılmalı ya da düzenleme yapılması zorunlu yorumları yapılmaktadır (Tekin, 2000; Turgut, 1992). Madde analizi sonucunda ayırt edicilik indisi 0.40 ve daha büyük olan maddeler ile .30 - .39 arasında olan maddeler testte bırakılmış, .20 - .29 aralığında olanlardan düzeltmelerle geliştirilemeyenler ve .19'dan küçük olanlar testten çıkarılmıştır.

Dördüncü aşama: Bu aşamada Webb (1997)'in son ölçütüne göre soruların kazanımlara dağılımının uygunluğu denge indeksi formülü ile hesaplanır.

3. Bulgular

Bu bölümde, geliştirilen başarı testine ilişkin geçerlik ve güvenilirlik analizlerinden elde edilen bulgular sunulmuştur.

Testin Kapsam Geçerlilik Çalışmasına İlişkin Bulgular

Testin kapsam geçerliğinin sağlanması için açıklanan test geliştirme aşamaları gerçekleştirilmiştir. Bunun için Webb (1997) tarafından ortaya konan ölçütler göz önüne alınarak uzman görüşleri alınmıştır. Alınan uzman görüşlerinin uyumluluğunun incelenmesi için Lawshe (1975) tekniğinden faydalanılmıştır. Kapsam geçerliğinin incelenmesi için Webb'in birinci ve üçüncü ölçütleri ele alınmıştır. Bunun için Ek-1'de verilen belirtke tablosuna uygun maddeler hazırlanmış ve uzman görüşüne sunulmuştur. Bu ölçütlere ilişkin bulgular Tablo 3 ve Tablo 4'te sunulmuştur. Tablo 3'de sunulduğu üzere .05 anlamlılık düzeyinde kapsam geçerlik oranlarının ortalaması ile elde edilen kapsam geçerlik indeksi 1 olarak bulunmuştur. Lawshe tekniği için tek boyutta 8 uzman görüşüne dayalı kapsam geçerlik oranı belirlenen ölçüt olan .78'in üzerinde olduğundan istatistiksel olarak anlamlı kabul edilmiştir. Buradan çıkarımla, testin tüm kazanımları kapsadığı sonucuna varılmıştır.

Tablo 3. Webb'in Birinci Kriterine Göre Kapsam Geçerlilik Ölçütleri

Kazanım Numarası	Yeterlidir	Yetersizdir	KGO
4.1.1	8	0	1
4.2.1	8	0	1
4.2.2	8	0	1
4.3.1	8	0	1
4.3.2	8	0	1
4.4.1	8	0	1
4.4.2	8	0	1
4.5.1	8	0	1
4.6.1	8	0	1
4.7.1	8	0	1
4.8.1	8	0	1

Uzman sayısı: 8, Kapsam geçerlik indeksi: 1

Tablo 4'te kazanımların sorularla uyumlu olmasının uygunluğunun değerlendirme sonuçları yer almaktadır. Buna göre, .05 anlamlılık düzeyinde $KGO \leq .78$ olan soruların yeniden düzenlenmesi ya da testten çıkarılması gerekmektedir. Çalışmada, 7. , 8. ve 21. sorular testten çıkartılmıştır. 2. , 4. , 5. , 9. , 11. ,12. ,13. ,16. , 18. , 21. , 22. , 23. ve 28. sorular uzman görüşleri doğrultusunda düzenlenerek teste alınmıştır. Sonuç olarak, 35 maddelik test 32 maddeye düşürülmüştür. Çıkarılan sorular sonucunda hesaplanan $KGİ \geq .78$ olduğundan testteki maddelerin kazanımlarla .05 düzeyinde istatistiksel olarak anlamlı olduğu görülmektedir.

Tablo 4. Webb'in İkinci Kriteri İle İlgili KGO Değerleri

Madde No	Uzman Sayısı	Uygun Dur	KGO	Madde No	Uzman Sayısı	Uygun Dur	KGO
1	8	8	1.00	19	8	8	1.00
2*	8	6	0.50	20	8	8	1.00
3	8	8	1.00	21**	8	5	0.25
4*	8	6	0.50	22*	8	7	0.75
5*	8	7	0.75	23*	8	6	0.50
6	8	8	1.00	24	8	8	1.00
7**	8	3	-0.25	25	8	8	1.00
8**	8	2	-0.50	26	8	8	1.00
9*	8	6	0.50	27	8	8	1.00
10	8	8	1.00	28*	8	6	0.50
11*	8	6	0.50	29	8	8	1.00
12*	8	7	0.75	30	8	8	1.00
13*	8	7	0.75	31	8	8	1.00
14	8	8	1.00	32	8	8	1.00
15	8	7	0.75	33	8	8	1.00
16*	8	6	0.50	34	8	8	1.00
17	8	8	1.00	35	8	8	1.00
18*	8	7	0.75				

* Düzenlenen maddeler; ** Çıkarılan maddeler

Testin Yapı Geçerlik Çalışmasına İlişkin Bulgular

Testin yapı geçerliğini sağlamak için, testte yer alan maddelerin analizi yapılmıştır. Test maddelerinin belirlenen ayırt edicilik ve güçlük düzeyleri Ek-2 'de verilmiştir.

Maddelerin ayırt edicilik indekslerine bakıldığında -.13 ve .76 arasında; güçlük düzeylerinde ise .25 ve .83 arasında değiştiği görülmektedir.

Tablo 5. Pilot Test Maddelerinin Ayırt Edicilik İndekslerine Göre Özeti

	Toplam
≥ 0.40	22
0.30-0.39	8
0.20-0.29	1
≤ 0.19	1

Ek-2’de yer verilen tablo incelendiğinde, testte yer alan 2, 4, 5, 7, 8, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 21, 25, 26, 27, 29, 30, 32 numaralı maddelerin .40 ve üzeri ayırt edicilik indeksine sahip olan ayırt ediciliği çok yüksek sorular; 3, 6, 9, 12, 22, 23, 24, 31 numaralı soruların .30 ve .39 değerleri aralığında ayırt edicilik indekslerine sahip olan iyi sorular olduğu görülmektedir. Bu sorularda herhangi bir düzenleme yapılmadan başarı testinde kullanılmasına karar verilmiştir. 1 numaralı sorunun .20 ayırt edicilik indeksi ile düzenlenmesi gereken soru ve 28 numaralı soru da -.13 ayırt edicilik indeksi ile çıkarılması gereken soru olması nedeniyle testten çıkarılmıştır. Ayrıca, çalışmada soruların kazanımlara dağılımları, Webb’in dördüncü ölçütü olarak verilen formül kullanılarak incelendiğinde kabul edilebilir oran olarak .90 bulunmuştur.

$$1 - ((|1/11 - 2/30| \times 4 + |1/11 - 3/30| \times 6 + |1/11 - 4/30| \times 1) / 2) = .90$$

Formülde 11 hazırlanan ünitenin içerdiği kazanım sayısını, 30 ünite için hazırlanan sorular üzerinde yapılan analizler sonucunda nihai formda yer alan soru sayısını, 2 kazanım başına düşen soru sayısını 4 ise 2 soru düşen kazanım sayısını, 3 kazanım başına düşen soru sayısını, 6 kazanım başına 3 soru düşen kazanım sayısını; 4 kazanım başına düşen soru sayısını, 1 ise kazanım başına düşen soru sayısının 4 olduğu kazanım sayısını ifade etmektedir.

Testin Güvenirliğine İlişkin Bulgular

Geliştirilen testin Pearson momentler çarpımı yarı güvenirlilik katsayısı .85 olarak bulunmuştur. Bu değer Spearman Brown güvenirlilik katsayısına göre düzeltilindiğinde güvenirlilik katsayısı .83 olarak belirlenmiştir. Bu da geliştirilen testin güvenilir olduğunu göstermektedir.

4. Sonuçlar

Başarı düzeyini belirlemek, programın etkililiğini değerlendirmek için kazanımlarla uyumlu yüksek geçerliğe ve güvenilirliğe sahip testlerin geliştirilmesi oldukça önemlidir. Bu çalışmada, Webb’in (1997) kazanımlar ve sınavlar arasındaki uyumu sağlamak için ortaya koyduğu ölçütler esas alınarak ilkökul dördüncü sınıf Maddeyi Tanıyalım Ünitesi’nde yer alan kazanımlara ve bu kazanımların bilişsel düzeylerine uygun geçerli ve güvenilir çoktan seçmeli bir test oluşturmak amaçlanmıştır. Bu araştırmada, amaca uygun olarak, çalışma sonucunda ilgili ünite için geçerli ve güvenilir 30 sorudan oluşan bir başarı testi geliştirilmiştir.

Başarı testlerinin, öğrencilerin akademik başarılarını ölçülmesine yönelik araştırmalar için bir araç olarak kullanıldığı düşünüldüğünde, bu testlerin kapsam ve yapı bakımından geçerli testler olması önemlidir. Alan yazında geliştirilen başarı testlerinde izlenen aşamalar bu çalışma ile benzerdir (Açıkgöz ve Karslı, 2015; Akbulut ve Çepni, 2013; Akdeniz, Öztürk ve Bakırcı, 2017; Ateş, 2004; Bakırcı, Artun ve Şenel, 2016; Bilgin, Ay ve Coşkun, 2013; Bozdoğan ve Kavcı, 2016; Bozkurt, 2010; Can Şen ve Eryılmaz, 2011; Can ve Yıldırım, 2017; Çalık ve Ayas, 2003; Gönen, Kocakaya ve Kocakaya, 2011; Korkmaz ve Konukaldı, 2015; Saraç, 2018; Şahin ve Ercan, 2015). Ancak, bu aşamalardan biri olan kapsam geçerliğinin sağlanması için uzman görüşlerinin alınması ve bu görüşler arasındaki uyumun geçerli olup olmadığının tespit edilmesi bakımından çalışma Akbulut ve Çepni’nin (2013) çalışması ile benzemekte ve bu yönüyle diğer çalışmalardan ayrılmaktadır.

Konu içeriği bakımından çalışmanın ele aldığı maddeyi tanıyalım ünitesine yönelik geliştirilen başarı testlerinin lise (Çalık ve Ayas, 2003) ve ortaokul (Ateş, 2004; Avcı, Acar Şeşen, Kırbaşlar, 2018; Bilgin, Ay ve Coşkun, 2013; Bozkurt, 2010; Saraç, 2018) düzeyinde olduğu görülmektedir. Değişen öğretim programı kazanımlarına uygun ilkökul dördüncü sınıf düzeyinde fen bilimleri dersi Maddeyi Tanıyalım Ünitesi için başarı testine rastlanmamıştır.

Başarı testi ve kavram testi gibi konuların kazanımlarını ele alan ölçme araçlarının geliştirilmesi süreci belirlenen basamaklara dikkat edilerek gerçekleştirilmelidir. Kazanımların incelenmesi, bu kazanımların bilişsel basamaklardan hangilerini kapsadığının belirlenmesi hazırlanacak sorunun kurgusunu oluşturma açısından oldukça önemlidir. Ayrıca

bu durum, öğrencilerin hangi bilişsel algı düzeyini ne ölçüde gerçekleştirebildiğini analiz etmede kullanılmaktadır. Soruların, kazanımların incelenmesi, alan yazının araştırılması sonucu oluşturulması ile birlikte test hazırlama süreci yeni bir aşamaya geçmektedir. Bu aşamada, uzmanlarca soruların kapsam geçerliklerinin, yaşa uygunluğunun, dil ve görsel açıdan anlaşılabilirliğinin incelenmesi gerekmektedir. Tüm bu incelemeleri takiben, hazırlanacak örnek bir test formu öğrencilere uygulanarak her maddenin gücü ve ayırt ediciliği hesaplanmalıdır. Bu hesaplamalar incelenerek ana test formunda yer alacak sorular belirlenmelidir. Bir testin geçerli, güvenilir ve kullanışlı olabilmesi belirtilen aşamaların titizlikle gerçekleştirilmesine bağlıdır.

5. Öneriler

Bu araştırmada, kazanım temelli hazırlanan çoktan seçmeli başarı testinin, öğrencilerin öğrenme düzeyleri hakkında detaylı bilgi vereceği düşünülmektedir. Geliştirilen testin, madde ünitesi başta olmak üzere, fen bilimleri dersinde ve diğer branşlarda geliştirilecek başarı testlerinin oluşturulma sürecine katkı sağlayacağına inanılmaktadır. Çalışmada yer alan kazanımlarla uyum ölçütleri, kapsam geçerliğinin sağlanması ve madde analizlerinin yapılması süreçleri, başarılı bir ölçme aracı geliştirmeye yardımcı olacaktır. Buradan çıkarımla, maddeyi tanıyalım ünitesinin öğretim programında ortaokul düzeyinde hazırlanacak testler için de örnek teşkil edebileceği düşünülmektedir. Ayrıca, araştırmacılar bu çalışmadaki test geliştirme süreçlerinden faydalanarak, diğer fen bilimleri konularına ve farklı alanlara yönelik geçerli ve güvenilir testler geliştirebilirler.

6. Kaynakça

- Açıköz, M. ve Karlı, F. (2015). Alternatif ölçme-değerlendirme teknikleri kullanılarak iş ve enerji konusunda geliştirilen başarı testinin geçerlilik ve güvenilirlik analizi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 4(1), 1-25.
- Akbulut, H. İ. ve Çepni, S. (2013). Bir üniteye yönelik başarı testi nasıl geliştirilir? İlköğretim 7. sınıf kuvvet ve hareket ünitesi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 2(1), 18-44.
- Akdeniz, A. R., Öztürk, M. ve Bakırcı, H. (2017). Bilgisayar destekli öğretim uygulamalarının sekizinci sınıf öğrencilerinin fen dersi akademik başarılarına ve bilginin kalıcılığına etkisi. *HAYEF: Journal of Education*, 14(2), 59-77.
- Ateş, M. (2004). *İşbirlikli öğrenme yönteminin ilköğretim II. kademede madde ve özellikleri ünitesinde öğrenci başarısına etkisi* (Yayımlanmamış yüksek lisans tezi). Dokuz Eylül Üniversitesi, İzmir.
- Avcı, F., Acar Şeşen, B. ve Kırbaslar, F.G. (2018). Maddenin yapısı ve özellikleri ünitesine yönelik iki aşamalı teşhis testinin geliştirilmesi. *Kastamonu Education Journal*, 26(4), 1007-1019. doi:10.24106/kefdergi.434239
- Ayhan, İ. (2010). Eğitimcilerde yol göstermesi açısından TAB analiz programı kullanarak başarı testi hazırlama sürecinde izlenecek adımlar. *Gümüşhane Üniversitesi Sosyal Bilimler Elektronik Dergisi*, 2, 79-101.
- Ayvacı, H. Ş. ve Türkdöğen, A. (2010). Yeniden yapılandırılan Bloom taksonomisine göre fen ve teknoloji dersi yazılı sorularının incelenmesi. *Türk Fen Eğitimi Dergisi*, 7(1), 13-25.
- Balcı, A. (2011). *Sosyal bilimlerde araştırma: Yöntem, teknik ve ilkeler (9.baskı)*. Ankara: Pegem Akademi.
- Bakırcı, H., Artun ve H. Şenel, S. (2016). Ortak bilgi yapılandırma modeline dayalı fen öğretiminin ortaokul yedinci sınıf öğrencilerinin kavramsal anlamalarına etkisi (gök cisimlerini tanıyalım). *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi (YYU Journal Of Education Faculty)*, 10 (1), 514-543
- Baykul, Y. (2015). *Eğitimde ve psikolojide ölçme: klasik test teorisi ve uygulaması (2. baskı)*. Ankara: Pegem Akademi
- Bilgin İ., Ay, Y. ve Coşkun, H. (2013). 5E modelinin ilköğretim dördüncü sınıf öğrencilerinin madde konusundaki başarılarına etkisinin ve model hakkındaki öğrenci görüşlerinin incelenmesi, *Kastamonu Eğitim Dergisi*, 21 (4), 1449-1470.
- Bozdoğan, E. A. ve Kavcı, A. (2016). Sınıf dışı öğretim etkinliklerinin ortaokul öğrencilerinin fen bilimleri dersindeki akademik başarılarına etkisi. *Gazi Eğitim Bilimleri Dergisi*, 2(1), 13-30.
- Bozkurt, E. (2010). *İlköğretim beşinci sınıf fen ve teknoloji dersi maddenin değişimi ve tanınması ünitesinde gazetelerden yararlanılarak hazırlanan ders etkinliklerinin tutum, başarı ve eleştirel düşünme becerilerine etkisi*. (Yayımlanmamış yüksek lisans tezi). Kocaeli Üniversitesi, Kocaeli.

- Brown, S. R. ve Conley, T.D. (2007). Comparing state high school assessments to standards success in entry-level university courses. *Educational Assessment*, 12(2), 137-160
- Can, S. ve Yıldırım, M. (2017). Eğitsel oyunlarla fen dersine var mısın yok musun?. *Atatürk Üniversitesi Kazım Karabekir Eğitim Fakültesi Dergisi*, 35, 13-29.
- Can Şen, H. ve Eryılmaz, A. (2011). Bir başarı testi geliştirme çalışması: basit elektrik devreleri başarı testi geçerlik ve güvenilirlik araştırması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 8 (1), 1-39.
- Çalık, M. ve Ayas, A. (2003). Çözeltilerde kavram başarı testi hazırlama ve uygulama. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 2(14), 1-17.
- Dindar, H. ve Demir, M. (2006). Beşinci sınıf öğretmenlerinin fen bilgisi dersi sınav sorularının Bloom taksonomisine göre değerlendirilmesi. *Gazi Eğitim Fakültesi Dergisi*, 26(3), 87-96.
- Edwards, N. (2010). An analysis of the alignment of the grade 12 physical sciences examination and the core curriculum in South Africa. *South Africa Journal of Education*, 30, 571-590.
- Ercan, S. ve Şahin, F. (2015). Fen eğitiminde mühendislik uygulamalarının kullanımı: tasarım temelli fen eğitiminin öğrencilerin akademik başarıları üzerine etkisi. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED)*, 9 (1), 128-164.
- Gönen, S., Kocakaya, S. ve Kocakaya, F. (2011). Dinamik konusunda geçerliliği ve güvenilirliği sağlanmış bir başarı testi geliştirme çalışması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 8(1), 40-57.
- Güler Demirci, P. M. (2008). Sınıf öğretmen adaylarının fen ve teknoloji dersinde öğrenilmesi gereken en önemli konunun ne olduğuna ilişkin düşünceleri. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 9(1), 113-121.
- Kara, Y. ve Çepni, S. (2011). Investigation the alignment between school learning and entrance examinations through item analysis. *Journal of Baltic Science Education*, 10(2), 73-86.
- Korkmaz, H. ve Konukaldı, I. (2015). İlköğretim fen ve teknoloji eğitiminde disiplinlerarası tematik öğretim yaklaşımının öğrencilerin öğrenme ürünleri üzerine etkisi. *Buca Eğitim Fakültesi Dergisi*, 39, 1-22.
- Kutlu, Ö., Yalçın, S. Ve Pehlivan, B.E. (2010). İlköğretim programında yer alan kazanımlara dayalı soru yazma ve puanlama çalışması. *İlköğretim Online*, 9 (3), 1201-1215.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.
- Özçelik, D. A. (2010). *Test hazırlama kılavuzu (4.baskı)*. Ankara: Pegem Akademi
- Pressley, M., Yokoi, L., Meter, P., Etten, V. S. ve Freebern, G. (1997). Some of reason why preparing for exams is so hard: what can be done to make it easier? *Educational Psychology Review*, 9(1), 1-38.
- Saraç, H. (2018). Fen Bilimleri Dersi 'Maddenin değişimi' ünitesi ile ilgili başarı testi geliştirme: Geçerlik ve güvenilirlik çalışması. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 18 (1), 416-445.
- Turgut, M. F. ve Baykul, Y. (2014). *Eğitimde ölçme ve değerlendirme (6.Baskı)*. Ankara: Pegem Akademi
- Türkiye İstatistik Kurumu (2013). *Seçilmiş göstergelerle İstanbul*. Ankara: TÜİK, ISSN 1307-0894, 97.
- Uyanık, G. (2014). *İlkokul dördüncü sınıf fen ve teknoloji dersinde kavramsal değişim yaklaşımının etkililiğinin incelenmesi* (Yayımlanmamış doktora tezi). Gazi Üniversitesi, Ankara.
- Veneziano L. ve Hooper J. (1997). A method for quantifying content validity of health-related questionnaires. *American Journal of Health Behavior*, 21(1), 67-70.
- Webb, N. L. (1997). *Determining alignment of expectations and assessments in mathematics and science education. NISE brief 1(2)*. Madison, WI: University of Wisconsin-Madison, National Institute for Science Education.
- Yan, X. ve Erduran, S. (2008). Arguing online: Case studies of pre-service science teachers' perceptions of online tools in supporting the learning of arguments. *Journal of Turkish Science Education*, 5(3), 2-31.
- Yurdugül, H. (2005). Ölçek geliştirme çalışmalarında kapsam geçerliği için kapsam geçerlik indekslerinin kullanılması. *XIV. Ulusal Eğitim Bilimleri Kongresi, Pamukkale Üniversitesi Eğitim Fakültesi, Denizli*.

EK-1. İlkokul Dördüncü Sınıf Maddeyi Tanıyalım Ünitesi Başarı Testi İçin Belirtke Tablosu

Kazanımlar	Bilişsel Basamak							Top.	%
	Bilgi	Kavrama	Uygulama	Analiz	Değerlendirme	Sentez			
4.3.1.1 Beş duyu organını kullanarak maddeyi niteleyen temel özellikleri açıklar.		*3						3	9
4.3.2.1 Maddenin hallerini bilir ve aynı maddenin farklı hallerine örnekler verir.	1	*2						3	9
4.3.2.2 Maddenin hallerine ait temel özelliklerini karşılaştırır.	1	*3						4	11
4.3.3.1 Farklı maddelerin kütle ve hacimlerini ölçerek karşılaştırır.	1	*	*2					3	9
4.3.3.2 Ölçülebilir özelliklerini kullanarak maddeyi tanımlar. Kütlesi ve hacmi olmayan olguların(ışık, ısı, gölge vb.) madde olmadığını belirtir.	*3							3	9
4.3.4.1 Maddenin ısınma-soğumasına yönelik deneyler tasarlar ve yapar.		3				*		3	9
4.3.4.2 Maddelerin ısı etkisiyle hal değiştirebileceğine yönelik deney yapar ve sonuçlarını yorumlar.	3				*	*		3	9
4.3.5 Madde ve cisim tanımlayarak aralarındaki farkları açıklar.	*1	*2						3	9
4.3.6.1 Günlük yaşamında sıklıkla kullandığı maddeleri saf madde, karışım şeklinde sınıflandırır ve aralarındaki farkları açıklar.	1	*2						3	9
4.3.7.1 Günlük yaşamda karşılaştığı karışımların ayrıştırılmasında kullanılacak yöntemlere karar verir ve test eder.	2	*2	*					4	11
4.3.8.1 Karışımları ayırmayı, ülke ekonomisine katkısı ve kaynakların etkili kullanımını bakımından önemini fark eder.	3				*			3	9

* =Bilişsel Basamak, sayılar= kazanımın bilişsel basamağı ile ilişkili soru sayısı

EK-2. Testin Pilot Çalışmasında Yer Alan Maddelerin Güçlük ve Ayırtedicilik İndeksleri

No	G	A	B	C	D	p	d
----	---	---	---	---	---	---	---

1*	Ü	0	14	9	1	.25	.20
	A	3	15	9	1		
2	Ü	0	0	0	23	.66	.59
	A	4	7	4	9		
3	Ü	3	0	21	0	.38	.35
	A	4	4	12	4		
4	Ü	0	0	23	0	.71	.63
	A	6	5	8	5		
5	Ü	0	18	0	5	.46	.53
	A	9	5	4	6		
6	Ü	0	5	0	18	.45	.37
	A	3	8	3	9		
7	Ü	0	0	21	0	.63	.68
	A	5	9	5	5		
8	Ü	0	19	4	0	.55	.52
	A	4	7	9	4		
9	Ü	15	0	5	4	.41	.38
	A	6	4	4	10		
10	Ü	24	0	0	0	.76	.67
	A	8	1	5	9		
11	Ü	0	1	5	17	.50	.41
	A	1	7	6	7		
12	Ü	10	3	3	8	.35	.32
	A	3	9	4	5		
13	Ü	1	22	1	0	.69	.68
	A	3	6	8	6		
14	Ü	0	0	0	24	.56	.42
	A	1	3	4	14		
15	Ü	19	3	1	1	.53	.48
	A	8	10	1	4		
16	Ü	1	1	21	1	.55	.43
	A	4	3	11	5		
17	Ü	20	1	1	1	.71	.73
	A	3	11	6	4		
18	Ü	19	1	1	1	.45	.40
	A	9	3	3	7		
19	Ü	3	0	16	5	.74	.54
	A	8	8	3	4		
20	Ü	21	0	1	0	.70	.64
	A	6	10	3	4		
21	Ü	1	0	0	22	.62	.56
	A	6	1	8	8		
22	Ü	19	1	3	0	.47	.31
	A	12	4	6	1		
23	Ü	10	13	1	0	.42	.31
	A	6	5	6	6		
24	Ü	4	11	4	5	.55	.36
	A	7	1	8	5		
25	Ü	0	22	0	0	.80	.76
	A	8	4	4	6		
26	Ü	0	3	0	19	.68	.57
	A	4	4	8	10		
27	Ü	23	0	0	0	.83	.76
	A	4	6	8	4		
28*	Ü	4	15	4	1	-.18	-.13
	A	9	1	6	5		
29	Ü	0	20	1	1	.66	.56
	A	10	6	3	1		
30	Ü	1	21	1	0	.66	.64
	A	1	6	6	7		
31	Ü	5	17	1	1	.36	.37
	A	6	8	3	5		
32	Ü	15	1	6	1	.57	.50
	A	3	5	9	5		

G:Grup, Ü: Üst grup, A: Alt grup, p: güçlük indeksi, d: ayırtedicilik indeksi