

Bilimin Doğası Hakkındaki Görüşlerin Değerlendirilmesinde Alternatif Bir Yöntem: Kısa Hikâyeler Yöntemi

Mahmut Polat¹ Mehmet Fatih Taşar²

Özet: Bu çalışmanın amacı fen bilgisi öğretmen adaylarının bilimin doğası hakkındaki görüşlerini değerlendirmek için kısa hikâye yönteminin etkililiğini ortaya koymaktır. Bu amaç doğrultusunda çalışmada kullanılan kısa hikâyelerin ve onları puanlamak için kullanılan rubriklerin geçerlik, güvenilirlik çalışmalarına yer verilmiştir. Çalışma, toplamda 169 öğretmen adayının katılımıyla pilot ve asıl uygulama olmak üzere iki aşamada gerçekleştirilmiştir. Pilot uygulama 2009 yılında doksan öğretmen adayıyla; asıl uygulama ise 2010 yılının bahar döneminde yetmiş dokuz öğretmen adayıyla Bilimin Doğası ve Tarihi dersinde yapılmıştır. Bilimin gelişimsellik, test edilebilirlik ve yaratıcılık boyutlarına ilişkin verileri kapsayan bu çalışmada; yapılan analizlerle kısa hikâyelerin (KH) geçerlik ve güvenilirliğine ilişkin olumlu sonuçlar elde edilmiştir. KH'lerin bilimin doğası hakkındaki görüşleri saptamada bir yöntem olarak kullanılabilceği yapılan analizlerle ortaya konulmuştur.

Anahtar Kelimeler: Bilimin Doğası, Ölçme- Değerlendirme, Kısa Hikâyeler, Geçerlik- Güvenirlik.

Abstract: *An Alternative Method For Assessing Views About The Nature of Science: Vignettes*

The aim of this study is to determine the effectiveness of the short stories (vignette) for assessing the pre service science teachers' views about the nature of science. For this purpose, the validity and reliability studies of vignettes and rubrics used in the research are given. The research comprised two stages: the pilot and the authentic application. The piloting was carried out in 2009 spring term with the participation of 90 pre service science teachers who had taken History and Nature of Science course while the main study included 79 pre service teachers in 2010 spring term. At the end of the analyses positive results were obtained as to the validity and reliability of vignettes. It was understood that vignettes can be used as a method of determining opinions on the nature of science.

Key Words: *Nature of Science, Measurement- Assessment, Vignettes, Validity, Reliability.*

Giriş

Günümüzde birçok ülkede etkisini gösteren fen eğitiminin amaçlarına ilişkin reform hareketlerinde bilimsel okur-yazarlık fen eğitiminin en önemli hedeflerinden biri olarak gösterilmektedir. Bilimsel okuryazarlığın önemli boyutlarından biri de bilimin doğasına ilişkin anlayıştır (MEB, 2005; National Research Council [NRC], 1996; American Association for the Advancement of Science [AAAS], 1993; National Science Teacher Association [NSTA], 1982). Bilimin doğası konusu son dönemdeki reformlar kapsamında popüler bir çalışma alanı olmuştur. Bu çalışmaların önemli bir kısmında fen öğretmenlerinin bilimin doğasını kavramalarına yönelik program geliştirmeye ve bilimin doğasına ilişkin görüşleri ölçmeye yönelik ölçek geliştirmeye odaklanılmaktadır (McComas, Clough ve Almazroa, 2000; Lederman, Abd-El-Khalick, Bell ve Schwartz, 2002).

Bilimin epistemoloji ve sosyolojisine ilişkin görüşleri değerlendirmek üzere yıllardır çeşitli ölçme araçları geliştirilmiştir (Aikenhead, 1988). Bununla birlikte bu araçlar genellikle öğrencilerin test maddelerini aracı geliştiren kişilerle aynı şekilde anlayıp yorumladıkları yanlış varsayımını içermektedir. Hiçbir belirsizliğin olmadığı şeklindeki bu varsayım fen eğitimi araştırmacıları için geleneksel bir problem olmuştur (Lederman & O'Malley, 1990; Aikenhead & Ryan, 1992). Örneğin Lederman ve O'Malley (1990) bilimin geçiciliği konusunda öğrenci görüşlerinin araştırmacı ve öğrenciler tarafından kullanılan dilin farklılığına bağlı olarak değişkenlik gösterebildiğini belirtmişlerdir.

Aikenhead (1988) öğrenci ve öğretmenlerin bilimin doğası hakkındaki görüşlerini değerlendirmede uygun ölçme aracını seçmek için bir takım yardımcı bilgilere yer vermiştir. Dört farklı cevaplama şeklindeki belirsizlik durumu dil problemi bakımından incelenmiştir. Öğrencilerin yazılı cevapları (likert tipi, paragraflar ve çok seçenekli) ve görüşmelerden elde edilen cevaplar arasındaki uyumsuzluk ölçülmüştür. Bu konuda ulaşılan sonuçları Aikenhead (1992) şöyle sıralamıştır:

1) Likert tipi cevaplar sadece öğrenci inançlarındaki bir tahmini ileri sürer ve doğru bir şekilde değerlendirme tahmininde bulunma şansı çok düşüktür. Belirsizlik sık sık % 80'ler seviyesine ulaşır.

¹ Yrd. Doç. Dr. Nevşehir Üniversitesi Eğitim Fakültesi

² Prof. Dr. Gazi Üniversitesi Eğitim Fakültesi

*Bilimin Doğası Hakkındaki Görüşlerin Değerlendirilmesinde Alternatif Bir
Yöntem: Kısa Hikâyeler Yöntemi*

2) Paragraf cevapları için belirsizlik yaklaşık olarak % 35 ile % 50 seviyeleri arasındadır. Bu likert tipi cevaplardan daha iyi bir durumdur. Buradaki belirsizlik bazı öğrencilerin eksik veya anlaşılabilir paragraflar yazma eğiliminde olması gerçeğinden çıkarılmıştır.

3) Yarı-yapılandırılmış görüşmeler tahminen en kolay anlaşılır ve doğru veri sunar fakat veriyi toplamak ve analiz etmek için çok fazla zamana ihtiyaç duyulmaktadır. Belirsizlik sadece yaklaşık % 5'tir.

4) Deneysel olarak elde edilmiş, çok seçenekli cevap durumunda belirsizlik % 15 ile % 20 seviyelerine kadar düşmektedir.

Bilim kavramının ve işlevinin son yüzyıldaki değişimi göz önünde bulundurulursa bu kavramlara ilişkin ölçme yaklaşımlarının da sabit kalması düşünülemez. Bilimin doğasının felsefi, sosyolojik ve tarihsel alandaki kavramsallaştırma değişimleri bilim eğitimi camiasının son yüzyıldaki bilimin doğası kavramını yansıtmaktadır (Abd-El-Khalick & Lederman, 2000). Fen okuryazarlığının bir boyutu olan bilimin doğasının anlaşılması için çabalar 1900'lü yılların başlangıcına dayandırılabilir (Lederman, 1992). O zamanlar bilimin doğasının anlaşılması bilimsel süreçlerin ve bilimsel yöntemlerin anlaşılması ile eşdeğerdi. Ders kitaplarında ve okul programlarında yer verilen bilimin özellikleri bilimsel bilgiyi mutlak doğru, nesnel, tümevarım yönteminin uygulanmasıyla gelişen, insan hayal gücü ve yaratıcılığından yoksun olarak tanımlayan, geleneksel görüşlere dayanmaktaydı (Haidar, 1999). 1960'larda bilimin doğasının amaçları araştırma, bilimsel süreç becerileri, gözlem, hipotez içerme, deney tasarlama ve verilerin yorumlanmasına dayanmaktaydı. 1980'lerle beraber bilimin doğasının tanımlanmasında psikolojik ve sosyolojik faktörler etkili olmaya başlamıştır. Bu faktörler teori temelli gözlemler, bilimsel açıklamaların gelişmesinde insan yaratıcılığının rolü, bilimsel organizasyonların sosyal yapısı ve bilimsel iddiaların değerlendirilmesinde sosyal söylemlerin rolüdür (Abd-El-Khalick & Lederman, 2000).

Tarihsel süreç içerisinde bilimin doğası tanımının, bilim kavramıyla paralel değiştiği anlaşılmaktadır. Bu nedenle özellikle 1980 li yıllardan önce oluşturulan ölçeklerin geçerliliği hakkında kuşku duymak mümkündür. Çünkü bilime yüklenen misyon ve tanım zamanla değişime uğramıştır. Ayrıca Aikenhead (1988) tarafından ortaya konulan ölçme yöntemindeki belirsizlik durumlarından dolayı bu ölçme araçlarının birçoğunun geçerliliği tartışmalıdır. Bu açıdan günümüzde daha çağdaş ve farklı ölçme değerlendirme araçlarına, yöntemlere ihtiyaç duyulduğu söylenebilir.

Bilimin doğası hakkında öğrenci ve öğretmenlerin görüşlerini değerlendirmek amacıyla son 50 yıl içerisinde 30'dan fazla ölçme aracı geliştirilmiştir ve bunların büyük bir çoğunluğu 1980 öncesine aittir (Bell, 2008). Ancak hazırlanan ölçeklerin büyük bir kısmı kişilerin görüşlerini tam olarak detaylı bir biçimde açıklamaya imkân vermemektedir (Taşar, 2006). Çalışmaların büyük bir kısmında kullanılan farklı ölçek ve teknikler örneklemedeki bireylerin konuya ilişkin kavram ve tutumlarını tam olarak detaylı ve özgür bir biçimde her hangi bir yönlendirmeye maruz bırakmaksızın açıklamalarına olanak sağlamaktan uzak olduğu görülmüştür.

Fen eğitimindeki köklü reform hareketlerinde, farklı ülke ve kültürlerde gerçekleştirilmiş olmasına rağmen, öne çıkan birkaç unsur bütün reform hareketlerinde varlığını göstermiştir. Bunlardan en önemlisi fen eğitiminin en temel amacı olarak bireysel farklılıkları ne olursa olsun bütün öğrencilerin fen okuryazarı bireyler olarak yetiştirilmesidir (MEB, 2005; National Research Council [NRC], 1996; American Association for the Advancement of Science [AAAS], 1993; National Science Teacher Association [NSTA], 1982). Bu bağlamda farklı ders ve eğitim seviyelerinde bireysel farklılıkların dikkate alınması öğrencilerin başarısı için zorunlu bir koşuldur. Bu ihtiyaç kendisini hem öğretim yöntemlerinde hem de ölçme değerlendirme yöntemlerinde hissettirmektedir. Çünkü bütün öğrenciler aynı bilişsel, duyuşsal, psiko-motor özelliklere sahip değildir. Her biri farklı bireysel niteliklere sahiptir. Bu nedenle mümkün oldukça her öğrenciye uygun olabilecek farklı ölçme değerlendirme tekniklerinin eğitim öğretim ortamına kazandırılması gereklidir.

Yukarıda sıralanan nedenlerle daha önce öğretim amaçlı olarak kullanılan kısa hikâyeler bu çalışmada örneklemin bilimin doğasına ilişkin görüşlerini ölçmek için kullanılacaktır. Kullanılan bu yöntemle örneklemin bilimin doğasına ilişkin görüşlerini daha detaylı ve açık uçlu bir biçimde açıklayabilmelerine imkân sağlanmıştır. Yapılan bu çalışmayla kısa hikâyelerin bu konuda bir

ölçme değerlendirme aracı olarak alternatif bir yöntem şeklinde alana kazandırılması amaçlanılmaktadır.

Çalışmanın Amacı

Bu çalışmanın amacı fen bilgisi öğretmen adaylarının bilimin doğası hakkındaki görüşlerini değerlendirmek için alternatif bir yöntem olarak kısa hikâyeye yönteminin bu alanda nasıl kullanılabileceğini ortaya koymaktır. Bu amaç doğrultusunda çalışmada kullanılan kısa hikâyelerin ve onları puanlamak için kullanılan rubriklerin geçerlik, güvenilirlik çalışmaları sıralanmıştır.

Çalışmanın Kapsamı

Bu çalışma, araştırmacı tarafından oluşturulan içerisinde bilimin doğasıyla ilgili özelliklerin ve kavramların bulunduğu sekiz kısa hikâyeyi kapsamaktadır. Araştırmacı tarafından kullanılan kısa hikâyelerde bilimin sadece gelişimsellik, test edilebilirlik, yaratıcılık boyutlarıyla ilgili verilere yer verilmiştir. Bu çalışma, çalışmaya katılan yüz altmış dokuz fen bilgisi öğretmen adayı ve bilimin doğası alanında çalışmış veya doktora yapmış beş uzmandan toplanan verileri kapsamaktadır.

Yöntem

Çalışma Grubu

Araştırmanın çalışma grubu, Ankara'da ki bir üniversitenin fen bilgisi öğretmenliği üçüncü sınıfında okuyan; pilot çalışma aşamasında 90 ve asıl uygulama aşamasında 79 kişi olmak üzere toplam 169 kişiden oluşmaktadır. Hem pilot uygulama hem de asıl uygulama, araştırmanın amacına uygun olarak zengin ve derinlemesine verilere ulaşmak için Bilimin Doğası ve Bilim Tarihi dersinden sonra gerçekleştirilmiştir.

Ayrıca araştırmada kısa hikâyelerin geçerlik ve güvenilirlik çalışmalarının farklı aşamalarında çeşitli üniversitelerden beş alan uzmanının görüş ve önerileri çalışmaya katılmıştır. Bu uzmanların tamamı doktorasını bilimin doğasıyla ilgili konularda yapmış veya bu konularda çalışmalar sunmuş kişilerden oluşmaktadır. Uzman grubun bu alandaki tecrübeleri beş ila yedi yıl arasında değişmektedir.

Veri Toplama Süreci

Bu çalışma pilot ve asıl uygulama olmak üzere iki aşamadan oluşmaktadır. Pilot uygulama aşamasında 2008- 2009 öğretim yılı bahar döneminde Bilimin Tarihi ve Doğası dersini alan 90 öğretmen adayına dönemin sonunda toplam 15 kısa hikâyeye (KH) uygulanmıştır. Bu hikâyeler içerisinde yer alan sorularla katılımcıların bilimin doğasına ilişkin görüşleri saptanmaya çalışılmıştır. Daha sonra amaçlı örneklem doğrultusunda seçilen 3 kişiyle yüz yüze görüşme gerçekleştirilerek bu uygulama hakkında daha detaylı görüşleri alınmıştır. Kısa hikâyelerde karşılaştıkları problemler, anlamadıkları bölümler tespit edilerek bu sorunların asıl uygulamada giderilmesine çalışılmıştır. Ayrıca hazırlanan örnek kısa hikâyeler konu alanında uzman olan kişilere sunulmuş ve onlardan gelen öneri ve uyarılarla bunlara son halleri verilmiştir.

Pilot çalışma sonrasında 79 fen bilgisi öğretmen adayından oluşan katılımcı gruba son şekli verilen sekiz kısa hikâyeye uygulanmıştır. Kısa hikâyeler öğretmen adaylarına 2009- 2010 öğretim yılının bahar döneminde Bilimin Doğası ve Bilim Tarihi dersinde on hafta ilgili öğretim elemanı dersi tamamladıktan sonra, araştırmacı tarafından uygulanmıştır.

Kısa hikâyelerden elde edilen yazılı cevapların değerlendirilmesinde rubrikler kullanılmıştır. Bu anlamda çalışmada veri toplama araçlarından bir diğerinin de rubrikler olduğu söylenebilir. Katılımcıların kısa hikâyelerdeki sorulara vermiş oldukları cevaplara göre rubriklerden bir puan seti oluşturulmuştur.

Veri Toplama Araçları

Kısa hikâyeler (KH) daha önceleri eğitimin farklı alanlarında farklı formlarda bilgi toplamak için kullanılmıştır. Kısa hikâyeler nitel olma özelliklerinden dolayı daha derinlemesine bilgi sağlamaları

ve küçük bir örnekleme uygulanabilmesi gibi belli avantajlara sahiptir (Taşar, 2006). Bu çalışmada kullanılan kısa hikâyeler Türkiye’de yayınlanmakta olan popüler dergilerden, kitaplardan ve gazete yazılarından derlenmiştir. Bu kısa hikâyeler içerisinde bilimin doğasına ilişkin unsurlar ve özellikler belli olmayan bir şekilde yer almaktadır. Her bir kısa hikâyede bilimin doğasına ilişkin tek bir özellik veya birden fazla özellikte yer alabilmektedir. Bir öykü akışı içerisinde olan kısa hikâyelerin bilim tarihi içerisinde önemli yer tutan paradigma dönüşümlerine öncülük eden bilim insanlarının yaşamlarından kesitler de içerebilmektedir.

Kısa hikâyelerin geçerliliği konu alanı uzmanlarının görüşlerine başvurularak sağlanılmıştır. Çalışmada kullanılan kısa hikâyelerin güvenilirliği için ise katılımcıların rubrik puanları arasındaki ilişki ve uyuma bakılmıştır. Böylece katılımcıların kısa hikâyelere vermiş oldukları cevapların değerlendirilmesinde kullanılan rubrik puanları arasındaki tutarlılığın ölçüsüyle güvenilirlik ortaya konulmaya çalışılmıştır.

Yapılan bu çalışmada kısa hikâyeler öğretmen adaylarının bilimin doğasına ilişkin görüşlerini belirleyebilmek için ölçme aracı olarak kullanılmıştır. Bunun için kısa hikâyelerde Rubba ve Andersen (1978) ’nin bilimsel bilginin doğası ölçeğinde yer alan sadece üç alt boyutuna (gelişimsellik, test edilebilirlik, yaratıcılığa) ilişkin sorular hazırlanmıştır. Literatürde bu boyutlarda hem çok fazla çalışma yapıldığı hem de bunlarda çok fazla kavram yanılgıları tespit edildiği için sadece bu boyutlar seçilmiştir. Kısa hikâyelerin ölçtüğü bilimin doğası alt boyutlarına göre dağılımı tablo 1’de verilmiştir. Aşağıda kısa hikâyelerin oluşturulma süreçleri aşamalarıyla açıklanmıştır:

1. Aşama: Araştırmacı çalışmanın başında çoğunluğu Tübitak yayınlarından olan birçok görsel ve yazılı materyali tarayarak çalışmanın amacına uygun olduğunu düşündüğü hikâyeleri toplamıştır. Daha sonra çalışmada uygulama pratikliği sağlaması için bu hikâyeler uygun bölümlerinden kesilerek parçalara ayrılmıştır. Böylece uzun bir hikâyeden bazen birden fazla kısa hikâye oluşturulmuştur (tablo 1). Kısa hikâyelerin bilimin doğası hakkında hangi boyut/boyutlarla ilişkili olabileceği belirlenmiştir. Muhtemel boyutlar belirlendikten sonra bu boyutlar hakkında derinlemesine bilgi almayı amaçlayan sorular her bir kısa hikâyenin altına yazılmıştır. Böylece ilk aşamanın sonunda araştırmacı tarafından toplam 10 uzun hikâyeden oluşturulan 21 kısa hikâye (vignetler) elde edilmiştir.

2. Aşama: Birinci aşamanın sonunda elde edilen 21 kısa hikâye araştırmacıyla beraber iki konu alanı uzmanı arasında tartışılmıştır. Her bir kısa hikâyenin sorularıyla birlikte incelendiği bu aşama sonunda altı kısa hikâye elenerek kısa hikâye sayısı 15’e düşürülmüştür. Elenen kısa hikâyelerin uygun kurguda olmaması, hikâyede yer alan konu ve kişilerin yeterince açık olmaması, çok uzun olması, ilgili bilimin doğası boyutunu ölçebilecek daha uygun özellikte kısa hikâyelerin var olması gibi nedenlerle elendiğini belirtmek gerekiyor.

3. Aşama: Bu aşamada elde var olan toplam 15 kısa hikâye ile pilot uygulamaya gidilmiştir. Toplam 90 katılımcıyla yapılan pilot uygulama sonrasında 3 kişiyle yarı yapılandırılmış görüşme gerçekleştirilmiştir. Bu görüşmede uygulanan kısa hikâyelere ilişkin sorular sorulmuştur. Ayrıca bu aşamada eldeki kısa hikâyeler farklı üniversitelerden olan beş konu alanı uzmanına bir protokolle birlikte gönderilerek onların da görüşleri alınmıştır. Hem pilot uygulamadan hem de beş uzmandan gelen dönütlere göre bu aşamada beş kısa hikâyenin daha elenmesinin uygun olacağı düşünülmüştür. Bu aşamada elenen kısa hikâyelerin diğerlerine göre okunması ve anlaşılması biraz daha zor olan hikâyeler olduğu hem pilot uygulamaya katılan öğretmen adaylarından hem de uzmanlardan gelen dönütlerden anlaşılmuştur.

4. Aşama: Üçüncü aşamanın sonunda uygulanmaya hazır 10 kısa hikâye varken araştırmacı nitel çalışmaların ruhuna da uygun olması nedeniyle başlangıçta ölçmeği düşündüğü altı boyuttan sadece üçünü ölçmeye karar vermiştir. Böylece bilimin doğası hakkında katılımcıların gelişimsellik, test edilebilirlik, yaratıcılık boyutları hakkında daha detaylı ve derinlemesine bilgileri toplayabileceğini düşünmüştür. Bu nedenle iki kısa hikâye daha çalışmadan elenerek sadece üç boyutu ölçen toplam 4 uzun hikâyeden oluşturulan 8 kısa hikâye ile asıl çalışmaya başlanılmıştır.

Bu çalışmada kullanılan sekiz kısa hikâye (vignet), büyük hikâyenin içinden uygun şekilde bölünerek oluşturulmuştur. Örneğin birinci, ikinci, üçüncü kısa hikâye “Yukarıda Ne var Ne Yok?”

adlı uzun hikâyeden oluşturulmuştur. Kısa hikâyelerde yer alan soruların bilimin doğası boyutlarına göre dağılımı tablo 1’de verilmiştir.

Tablo 1. Kısa Hikâyelerdeki Soruların Bilimin Doğası Boyutlarına Göre Dağılımı

KH. No	GELİŞİMSELLİK	T. EDİLEBİLİRLİK.	YARATICILIK	TOPLAM
KH 1	1., 4. sorular	2. soru	3. soru	4 soru
KH 2	3., 4. sorular	1. 2. soru	-	4 soru
KH 3	1. 2. 3. 5. sorular	4. soru	-	5 soru
KH 6	1. soru	3. soru	-	2 soru
KH 7	1. soru	2. soru	-	2 soru
KH 8	1. soru	2. soru	-	2 soru
KH 10	3. 4. sorular	1. soru	2. 5. 6. sorular	6 soru
KH 12	-	-	1. 2. 3. sorular	3 soru
TOP.	13 soru	8 soru	7 soru	28 soru

Tablo 1’de görüldüğü gibi en fazla soru onuncu kısa hikâye içinden sorulmuştur. Çalışmanın tamamında gelişimsellik boyutu toplamda on üç, test edilebilirlik boyutu sekiz, yaratıcılık boyutu ise yedi soruyla değerlendirilmiştir. Kısa hikâyelerin değerlendirilmesinde geliştirilen rubrikler kullanıldığı için aşağıda rubriklere ilişkin bazı açıklamalara yer verilmiştir.

Bu çalışmada öğretmen adaylarının bilimin doğası hakkındaki görüşlerinin belirlenebilmesi için sunulan kısa hikâyelerde ki sorulara verdikleri cevaplar araştırmacı tarafından geliştirilen rubriklerle değerlendirilmiştir. Bodur (2003) rubriklerin geliştirilmesinde iki yol olduğundan bahsetmektedir. Bunların ilki literatürde yapılmış benzer çalışmaların uyarlanarak kullanılması diğeri ise katılımcıların vermiş oldukları cevaplardan derlenen ifadelerin konuya uygun bir şekilde düzenlenerek kullanılmasıdır. Hali hazırdaki bu çalışmada ise sayılan iki yolda kullanılarak ilk rubrik taslakları oluşturulmuştur. Rubriklerin hazırlanması aşamasında öncelikle literatürden (Abd-El-Khalick, Waters, ve Le, 2008; Özgelen, 2010) bilimin doğasıyla ilgili kullanılan rubrikler ve değerlendirme ifadeleri incelenmiştir. Söz konusu bu çalışmalarda kullanılan ölçütler ve nitelik ifadelerinden bu çalışmaya uygun olanlar seçilmiştir. Ayrıca çalışmaya katılan öğretmen adaylarının kısa hikâyelerde kullandıkları ifadelerden de uygun olanlar seçilerek rubriklerin taslakları oluşturulmuştur. Rubriklerin taslakları toplamda üç başarı seviyesi tanımlanarak gelişimsellik, test edilebilirlik ve yaratıcılık boyutları için ayrı ayrı hazırlanmıştır. Çalışmanın başında bu başarı seviyeleri ‘mükemmel / gerçekçi / çağdaş’ (5 puan), ‘orta/ makul’ (3 puan), ‘naif / geleneksel’ (1 puan) şeklinde sıralanmıştı, ancak rubriklerin geçerlik ve güvenilirlik çalışmaları sırasında uzmanlardan gelen öneriler dikkate alınarak başarı seviyelerine bir tane daha eklenmiştir. Eklenen bu seviye sorulara hiç cevap vermeyen ya da cevabı soruyla tamamen ilişkisiz olan katılımcılar için konulan ‘başarısız’ (0 puan) boyutudur.

Rubriklerin geçerlik çalışması için rubrik taslakları konu alanı uzmanlarına katılımcıların birinci ve onuncu kısa hikâye cevaplarıyla (45 katılımcının) birlikte gönderilmiştir. Uzmanlardan kısa hikâyelerdeki sorulara verilen cevaplarla rubrikteki ifadelerin karşılaştırılması istenmiştir. Rubriklerin bu cevapları değerlendirme de yeterli olup olmadığı sorulmuştur. Konu alanı uzmanlarından gelen değerlendirme ve önerilerle taslaklarda gerekli düzenlemeler yapılmıştır. Örneğin bazı yazılmış ifadelerin daha somut hale getirilmesi için katılımcıların vermiş oldukları

cümlelerden alıntılar yapılarak rubrikteki ifadelerin daha anlaşılır ve somut kriterlere kavuşması sağlanılmaya çalışılmıştır. Rubrikler bu düzenlemeler yapıldıktan sonra uzmanlara tekrar gönderilmiştir. Uzmanların olumlu görüşleri alındıktan sonra rubrikler çalışmada kullanılmıştır.

Rubriklerin güvenilirlik çalışması her ne kadar geçerlik çalışmalarıyla paralel bir şekilde yürütülmüşse de burada ayrı ayrı aşamalar halinde açıklanması uygun görülmüştür. Rubriklerin güvenilirliği için literatürde sıklıkla puanlayıcılar arası güvenilirlik katsayılarının (inter- rater reliability) hesaplandığı görülmüştür (Bodur, 2003). Bunun için iki konu alanı uzmanına gönderilen rubrikteki ifadelerden herkesin aynı durumları algılayıp algılamadıklarını belirleyebilmek için katılımcılara ait birinci ve onuncu kısa hikâyenin cevapları ve oluşturulan rubrikler birlikte gönderilmiştir. Uzmanlar katılımcılara ait cevapları okuyarak gelişimsellik, test edilebilirlik, yaratıcılık boyutuyla ilgili her birine bir puan vermişlerdir. Birinci ve onuncu kısa hikaye cevapları için bu işlem bitirildikten sonra aynı cevaplara araştırmacı da bir puan vermiştir. Böylece aynı katılımcılara ait belli cevaplara farklı uzmanların vermiş oldukları üç puan seti üzerinden güvenilirlik analizi yapılmıştır. Rubriklerin güvenilirlik analizi için uzmanların (puanlayıcıların) puan setleri arasındaki pearson korelasyon katsayısının hesaplanmasının uygun olduğu görülmüştür. Aşağıda bu analiz sonucunda elde edilen sonuçlar tablolar halinde sunulmuştur.

Tablo 2 Birinci Kısa Hikâye İçin Puanlayıcılar Arası Güvenirlik Sonuçları

Puanlayıcılar	Gelişimsellik	T. Edilebilirlik	Yaratıcılık
P ₁ – P ₂	0,83	0,87	0,81
P ₁ – P ₃	0,85	0,85	0,88
P ₂ – P ₃	0,77	0,81	0,77

Tablo 3 Onuncu Kısa Hikâye İçin Puanlayıcılar Arası Güvenirlik Sonuçları

Puanlayıcılar	Gelişimsellik	T. Edilebilirlik	Yaratıcılık
P ₁ – P ₂	0,92	0,89	0,75
P ₁ – P ₃	0,83	0,81	0,79
P ₂ – P ₃	0,75	0,78	0,77

Sonuç olarak hem uzmanlar arasında hem de uzmanlarla araştırmacı arasında hesaplanan puanlayıcılar arası güvenilirlik (inter- rater reliability) katsayısı 0,70'ten büyük çıkmıştır. Bodur'a (2003) göre bu sonuç rubriklerin güvenilirliği için yeterlidir. Buradan hareketle rubriklerin herkes tarafından aynı şekilde algılandığı ve bunlarla verilen puanların puanı veren kişilerden bağımsız olduğu söylenebilir. Özetle çalışmada kullanılan rubriklerin güvenilir oldukları sonucu çıkarılabilir.

Verilerin Analizi

Geçerlilik ve güvenilirlik çalışmaları yapılmış olan rubriklere göre araştırmacı tarafından katılımcıların her bir kısa hikâyede her bir boyuta ilişkin görüşleri tespit edilerek bu görüşlere bir puan takdir edilmiştir.

Bu çalışmanın amacına uygun olarak kullanılan kısa hikâyelerin güvenilirliğine ilişkin verilere ulaşabilmek için rubriklerden elde edilen nicel verilerin kendi içinde ki ilişkileri ve uyumu ortaya konulmuştur. Bunun için rubrik puanları Wilcoxon uyumlu çiftler- işaretli sıralar testi kullanılarak analiz edilmiştir. Bu testin dayandığı iki temel varsayım olarak puanların elde edildiği grupların ilişkili olması ve puanların en azından sıralama ölçeğinden elde edilmiş olması sayılabilir

(Büyüköztürk, ve diğ., 2009). Rubriklerden elde edilen ilişkili puan setleri arasında fark olup olmadığını ya da bir başka ifadeyle aralarındaki ilişkiyi ortaya koymak için gelişimsellik ve test edilebilirlik boyutlarında birinci, ikinci, üçüncü kısa hikâye puanları kendi arasında ve altıncı, yedinci, sekizinci kısa hikâye puanları yine kendi arasında değerlendirilerek analiz edilmiştir. Yaratıcılık boyutu için ise birinci, onuncu, on ikinci kısa hikâyeden elde edilen rubrik puanları kendi arasında analiz edilmiştir.

Katılımcılara uygulanan toplam sekiz kısa hikâye içerisindeki sorulara verilen cevaplar, kullandıkları ifadeler onların konu hakkındaki görüşlerini belirlemek için çok ciddi bir veri kaynağı oluşturmuştur. Ayrıca kısa hikâyelerin aynı boyutları ölçtüğü düşünüldüğünde aynı kişilerin belli bir konu hakkındaki görüşleri tekrarlı bir şekilde ölçüldüğü kabul edilebilir. Örneğin, gelişimsellik boyutu çalışma süresince yedi defa farklı kısa hikâyede değerlendirilmiştir. Böylece aynı öğretmen adayına ait tekrarlı ölçümler yapılmış olmaktadır.

Bulgular ve Yorum

Kısa Hikâyelerin Geçerliliğine Ait bulgular

Bir ölçme aracının ölçmek istenilen özelliği ölçme derecesine onun bu özelliği ölçmedeki geçerliği denir (Özçelik, 1997). Terminolojide farklı geçerlik türlerine rastlansa da çoğunlukla bilim adamlarınca kabul gören geçerlik türleri kapsam geçerliği, yordama geçerliği, yapı geçerliği olarak sıralanabilir (Özçelik, 1997).

Bu çalışmada kullanılan kısa hikâyelerin geçerliği için kapsam ve yordama geçerliğine bakılmıştır.

Kısa Hikâyelerin Kapsam Geçerliği

Çalışmada kullanılan kısa hikâyelerin kapsam geçerliği için hikâyelerin oluşturulması sürecinde alanda uzman olan beş öğretim elemanının görüşlerinden yararlanılmıştır. Kısa hikâyelerin oluşturulma süreci yöntem kısmında detaylı bir şekilde açıklanmıştır. Uzmanlardan alınan görüşlere göre çalışmada kullanılan kısa hikâyelerin çalışmanın amacı doğrultusunda kapsam geçerliğine sahip oldukları görülmüştür. Hem hikâyelerde anlatılan konular hem de bunlara ilişkin soruların bilimin doğası hakkındaki kavram ve ilkeleri kapsadığı sonucuna varılmıştır. Ayrıca bu durum katılımcıların tamamının hikâyelerden alınan puanlarında da görülmektedir. Çalışmaya katılan öğretmen adaylarının bütün rubrik puanları (toplam 1659 farklı rubrik puanı) incelendiğinde sorulan sorulara verilen cevapların konuyla ilgisizliğini gösteren sıfır (0) puanını sadece 35 defa aldıkları görülmüştür. Bu bulgudan hareketle de kullanılan kısa hikâyelerin kapsam geçerliği açısından yeterli oldukları söylenebilir.

Kısa Hikâyelerin Yordama Geçerliği

Yordama geçerliği; testlerden elde edilen bir “yordayıcı puan” ile gelecekteki durumlarla ilgili bir “ölçüt”e ilişkin değerler arasındaki korelasyon katsayısı olarak nitelendirilebilir. Bu çalışmada kullanılan kısa hikâyelerin yordama geçerliğini bulmak için katılımcıların KHY ile elde etmiş oldukları rubrik puanlarının toplamı ile dönem sonunda Bilimin Tarihi ve Doğası dersinden almış oldukları ders notu arasındaki korelasyon hesaplanmıştır. Böylece bir ölçme değerlendirme aracı olacak olan kısa hikâyelerin katılımcıların ilgili dersten almış oldukları ders notlarını yordama gücü saptanmaya çalışılmıştır. Katılımcıların Bilimin Tarihi ve Doğası dersi için ilgili öğretim elemanı tarafından verilen bir vize bir de final sınav puanı kullanılarak dönem sonu notları oluşturulmuştur.

Katılımcıların bütün rubriklerden almış oldukları toplam puanları ile dönem sonu ders notları arasında Pearson korelasyon katsayısı 0,86 olarak hesaplanmıştır. Hesaplanan katsayı 0.70’ten büyük çıkmıştır. Bu bulguya göre kısa hikâyelerin yordama geçerliği açısından yeterli olduğu söylenebilir.

Kısa Hikâyelerin Güvenirliğine Ait bulgular

Bir ölçme aracı hangi özelliği ölçüyorsa onun bu özelliğin gerçek değerlerine yakın ölçüler verdiğini savunabilmek için bir kere bu araç, özelliği ölçülen varlık ya da olayların bu özelliğinde bir değişme olmadıkça onları hep aynı sıraya koyan ölçüler vermelidir. *Ölçme aracı neyi ölçüyorsa onu, kararlı bir şekilde ölçmelidir. Buna ölçme aracının güvenilirliği denir* (Özçelik, 1997). Başka bir deyişle

güvenirlik; bir testin, ölçülmesi gereken şeyi her uygulamasında aynı biçimde ölçmede gösterdiği tutarlık derecesidir.

Bu çalışmada kullanılan kısa hikâyelerin güvenilirliğinin hesaplanabilmesi için kısa hikâyelerin değerlendirilmesinde kullanılan rubrik puanları arasındaki ilişkiye ve kararlılığa bakılması uygun görülmüştür. Çünkü elde edilen rubrik puanları bilimin ölçülen her bir boyutunu birden fazla defa tekrarlı bir biçimde değerlendirmiştir. Böylece bir öğretmen adayının aynı konu hakkındaki tekrarlı ölçümleri güvenilirliğinin hesaplanmasında önemli rol almıştır. Aşağıda sırayla gelişimsellik, test edilebilirlik, yaratıcılık boyutlarından alınan rubrik puanları arasındaki uyuma ve kararlılığa ilişkin bulgular sunulmuştur.

Tablo 4. Öğretmen adaylarının gelişimsellik ile ilgili birinci, ikinci ve üçüncü kısa hikâyeden almış oldukları rubrik puanlarına ait bulgular

Rubrik puanı	K. Hikâye 1		K. Hikâye 2		K. Hikâye 3	
	f	%	f	%	f	%
0	-	-	-	-	9	11,4
1	47	59,5	45	57	41	51,9
3	30	38	31	39,2	28	35,4
5	2	2,5	3	3,8	1	1,3
M	1,86		1,94		1,65	

Tablo 4'e göre her üç kısa hikâyede öğretmen adaylarının gelişimsellik ile ilgili almış oldukları rubrik puan ortalamaları için altında kalmaktadır. Katılımcıların yarısından fazlasının her üç hikâyeden de bir puan aldıkları görülmüştür. Kısa hikâyelerden tam puan alanların sayısı çok küçük bir yüzdeye denk gelmektedir.

Tablo 5. Öğretmen adaylarının gelişimsellik ile ilgili birinci, ikinci ve üçüncü kısa hikâyeden almış oldukları rubrik puanlarının Wilcoxon işaretli sıralar testi sonuçları

Hikâye Çiftleri		n	Sıra ort.	Sıra topl	z	P
Hikâye 1 – Hikâye 2	Negatif sıra	5	7	35	-0,832	0.40
	Pozitif sıra	8	7	56		
	Eşit	66	-	-		
Hikâye 1 – Hikâye 3	Negatif sıra	21	13,79	289,50	-1,220	0.22
	Pozitif sıra	9	19,50	175,50		
	Eşit	49	-	-		
Hikâye 2 – Hikâye 3	Negatif sıra	24	16,46	395	-1,75	0.08
	Pozitif sıra	10	20	200		

Eşit	45	-	-
------	----	---	---

p>0.05

Tablo 5'e göre öğretmen adaylarının gelişimsellik boyutuyla ilgili olarak birinci, ikinci, üçüncü kısa hikâye puanları arasında istatistiksel anlamda bir fark bulunamamıştır. Katılımcıların almış oldukları puan ortalamaları birbirine çok yakındır. Bu bulgudan hareketle öğretmen adaylarının gelişimsellik boyutuyla ilgili görüşlerini ortaya çıkarmak için kullanılan birinci, ikinci, üçüncü kısa hikâyenin istikrarlı bir şekilde bu işlevi gerçekleştirebildiği söylenebilir.

Tablo 6. Öğretmen adaylarının gelişimsellik ile ilgili altıncı, yedinci, sekizinci kısa hikâyeden almış oldukları rubrik puanlarına ait bulgular

Rubrik puanı	K. Hikâye 6		K. Hikâye 7		K. Hikâye 8	
	f	%	f	%	f	%
0	3	3,8	3	3,8	3	3,8
1	62	78,5	56	70,9	56	70,9
3	12	15,2	19	24,1	15	19
5	2	2,5	1	1,3	5	6,3
M	1,37		1,49		1,59	

Tablo 6'ya göre her üç kısa hikâye de öğretmen adaylarının gelişimsellik ile ilgili almış oldukları rubrik puan ortalamaları için altında kalmaktadır. Katılımcıların yarısından fazlasının her üç kısa hikâyeden de bir puan aldıkları görülmüştür. Kısa hikâyelerden tam puan alanların sayısı çok küçük bir yüzdeye denk gelmektedir.

Tablo 7. Öğretmen adaylarının gelişimsellik ile ilgili altıncı, yedinci, sekizinci kısa hikâyeden almış oldukları rubrik puanlarının Wilcoxon işaretli sıralar testi sonuçları

Hikâye Çiftleri		n	Sıra ort.	Sıra topl	z	P
Hikâye 6 – Hikâye 7	Negatif sıra	3	6	18	-1,51	0,13
	Pozitif sıra	8	6	48		
	Eşit	68	-	-		
Hikâye 6 – Hikâye 8	Negatif sıra	5	7,50	37,50	-1,95	0,051
	Pozitif sıra	12	9,63	115,50		
	Eşit	62	-	-		
Hikâye 7 – Hikâye 8	Negatif sıra	9	8,61	77,50	-0,74	0,46
	Pozitif sıra	10	11,25	112,50		
	Eşit	60	-	-		

p>0.05

Tablo 7'ye göre öğretmen adaylarının gelişimsellik boyutuyla ilgili olarak altıncı, yedinci, sekizinci kısa hikâye puanları arasında istatistiksel anlamda bir fark bulunamamıştır. Katılımcıların almış oldukları puan ortalamaları birbirine çok yakındır. Bu bulgudan hareketle öğretmen adaylarının gelişimsellik boyutuyla ilgili görüşlerini ortaya çıkarmak için kullanılan altıncı, yedinci, sekizinci kısa hikâyelerin istikrarlı bir şekilde bu işlevi gerçekleştirebildiği söylenebilir.

Tablo 8. Öğretmen adaylarının test edilebilirlik boyutuyla ilgili birinci, ikinci, üçüncü kısa hikâyeden almış oldukları rubrik puanlarına ait bulgular

Rubrik puanı	K. Hikâye 1		K. Hikâye 2		K. Hikâye 3	
	f	%	f	%	f	%
0	-	-	-	-	-	-
1	33	41,8	35	44,3	37	46,8
3	44	55,7	39	49,4	41	51,9
5	2	2,5	5	6,3	1	1,3
M	2,22		2,24		2,09	

Tablo 8'e göre her üç kısa hikâye de öğretmen adaylarının test edilebilirlik boyutuyla ilgili almış oldukları rubrik puan ortalamaları için altında kalmaktadır. Katılımcıların yarısına yakın bir kısmının her üç kısa hikâyeden de bir puan aldıkları görülmüştür. Kısa hikâyelerden tam puan alanların sayısı çok küçük bir yüzdeye denk gelmektedir.

Tablo 9 Öğretmen adaylarının test edilebilirlikle ilgili birinci, ikinci, üçüncü kısa hikâyeden almış oldukları rubrik puanlarının Wilcoxon işaretli sıralar testi sonuçları

Hikâye Çiftleri		n	Sıra ort.	Sıra topl	z	P
Hikâye 1 – Hikâye 2	Negatif sıra	4	5	20	-0,33	0,74
	Pozitif sıra	5	5	25		
	Eşit	70	-	-		
Hikâye 1 – Hikâye 3	Negatif sıra	7	5	35	-1,68	0,10
	Pozitif sıra	2	5	10		
	Eşit	70	-	-		
Hikâye 2 – Hikâye 3	Negatif sıra	9	7,22	65	-1,50	0,13
	Pozitif sıra	4	6,50	26		
	Eşit	66	-	-		

p>0.05

Tablo 9'a göre öğretmen adaylarının test edilebilirlik boyutuyla ilgili olarak birinci, ikinci, üçüncü kısa hikâye puanları arasında istatistiksel anlamda bir fark bulunamamıştır. Katılımcıların almış oldukları puan ortalamaları birbirine çok yakındır. Bu bulgudan hareketle öğretmen adaylarının test edilebilirlik boyutuyla ilgili görüşlerini ortaya çıkarmak için kullanılan birinci, ikinci, üçüncü kısa hikâyenin istikrarlı bir şekilde bu işlevi gerçekleştirebildiği söylenebilir.

Tablo 10. Öğretmen adaylarının test edilebilirlik boyutuyla ilgili altıncı, yedinci, sekizinci kısa hikâyeden almış oldukları rubrik puanlarına ait bulgular

Rubrik puanı	K. Hikâye 6		K. Hikâye 7		K. Hikâye 8	
	f	%	f	%	f	%

Bilimin Doğası Hakkındaki Görüşlerin Değerlendirilmesinde Alternatif Bir Yöntem: Kısa Hikâyeler Yöntemi

0	16	20,3	16	20,3	1	1,3
1	40	50,6	41	51,9	60	75,9
3	22	27,8	17	21,5	15	19
5	1	1,3	5	6,3	3	3,8
M	1,41		1,48		1,52	

Tablo 10'a göre her üç kısa hikâyede de öğretmen adaylarının test edilebilirlikle ilgili almış oldukları rubrik puan ortalamaları için altında kalmaktadır. Katılımcıların yarısından fazlasının her üç kısa hikâyeden de bir puan aldıkları görülmüştür. Kısa hikâyelerden tam puan alanların sayısı çok küçük bir yüzdeye denk gelmektedir.

Tablo 11. Öğretmen adaylarının test edilebilirlikle ilgili altıncı, yedinci, sekizinci kısa hikâyeden almış oldukları rubrik puanlarının Wilcoxon işaretli sıralar testi sonuçları

Hikâye Çiftleri		<i>n</i>	<i>Sıra ort.</i>	<i>Sıra topl</i>	<i>z</i>	<i>P</i>
Hikâye 6 – Hikâye 7	Negatif sıra	6	8	48	-0,77	0,44
	Pozitif sıra	9	8	72		
	Eşit	64	-	-		
Hikâye 6 – Hikâye 8	Negatif sıra	16	26,72	427,50	-0,31	0,76
	Pozitif sıra	26	18,29	475,50		
	Eşit	37	-	-		
Hikâye 7 – Hikâye 8	Negatif sıra	13	25,58	332,50	-0,01	0,99
	Pozitif sıra	23	14,50	333,50		
	Eşit	43	-	-		

$p > 0.05$

Tablo 11'e göre öğretmen adaylarının test edilebilirlik boyutuyla ilgili olarak altıncı, yedinci, sekizinci kısa hikâye puanları arasında istatistiksel anlamda bir fark bulunamamıştır. Katılımcıların almış oldukları puan ortalamaları birbirine çok yakındır. Bu bulgudan hareketle öğretmen adaylarının test edilebilirlik boyutuyla ilgili görüşlerini ortaya çıkarmak için kullanılan altıncı, yedinci, sekizinci kısa hikâyenin istikrarlı bir şekilde bu işlevi gerçekleştirebildiği söylenebilir.

Tablo 12. Öğretmen adaylarının yaratıcılık boyutuyla ilgili birinci, onuncu, on ikinci kısa hikâyeden almış oldukları rubrik puanlarına ait bulgular

Rubrik puanı	K. Hikâye 1		K. Hikâye 10		K. Hikâye 12	
	f	%	f	%	f	%
0	-	-	-	-	-	-
1	36	45,6	33	41,8	35	44,3
3	41	51,9	42	53,2	39	49,4
5	2	2,5	4	5,1	5	6,3
M	2,14		2,27		2,24	

Tablo 12'ye göre birinci, onuncu, on ikinci kısa hikâyede öğretmen adaylarının yaratıcılıkla ilgili almış oldukları rubrik puan ortalamaları için altında kalmaktadır. Katılımcıların yarısına yakın bir bölümünün her üç kısa hikâyeden de bir puan aldıkları görülmüştür. Kısa hikâyelerden tam puan alanların sayısı çok küçük bir yüzdeye denk gelmektedir.

Tablo 13. Öğretmen adaylarının test edilebilirlikle ilgili birinci, onuncu, on ikinci kısa hikâyeden almış oldukları rubrik puanlarının Wilcoxon işaretli sıralar testi sonuçları

Hikâye Çiftleri		<i>n</i>	<i>Sıra ort.</i>	<i>Sıra topl</i>	<i>z</i>	<i>P</i>
Hikâye 1 – Hikâye 10	Negatif sıra	1	4	4	-0,89	0,06
	Pozitif sıra	6	4	24		
	Eşit	72	-	-		
Hikâye 1 – Hikâye 12	Negatif sıra	1	3,50	3,50	-1,63	0,10
	Pozitif sıra	5	3,50	17,50		
	Eşit	73	-	-		
Hikâye 10 – Hikâye 12	Negatif sıra	4	4	16	-0,38	0,70
	Pozitif sıra	3	4	12		
	Eşit	72	-	-		

$p > 0.05$

Tablo 13'e göre öğretmen adaylarının yaratıcılık boyutuyla ilgili olarak birinci, onuncu, on ikinci kısa hikâye puanları arasında istatistiksel anlamda bir fark bulunamamıştır. Katılımcıların almış oldukları puan ortalamaları birbirine çok yakındır. Bu bulgudan hareketle öğretmen adaylarının yaratıcılık boyutuyla ilgili görüşlerini ortaya çıkarmak için kullanılan birinci, onuncu, on ikinci kısa hikâyenin istikrarlı bir şekilde bu işlevi gerçekleştirebildiği söylenebilir.

Öğretmen adaylarının bilimin doğası hakkındaki görüşlerini belirlemek için kısa hikâyelerin kullanıldığı bu çalışmada kısa hikâyelere verilen cevapların değerlendirildiği rubrik puanları önem kazanmaktadır. Gelişimsellik boyutunda gerek birinci, ikinci, üçüncü kısa hikâyeden elde edilen bulgular (tablo 4 ve 5); gerekse altıncı, yedinci, sekizinci kısa hikâyeden elde edilen bulgular (tablo 6 ve 7) birlikte değerlendirildiğinde kısa hikâyelerin katılımcıların bu boyuta ilişkin görüşlerini ölçme ve değerlendirme de güvenilirlik özelliğine sahip olduğu söylenebilir. Aynı şey hem test edilebilirlik boyutu için (tablo 8, 9, 10, 11) hem de yaratıcılık boyutu için (tablo 12 ve 13) söylenebilir.

Ayrıca kısa hikâyelere ait betimsel istatistiklerin sunulduğu tablolar (tablo 4, 6, 8, 10, 12) incelendiğinde öğretmen adaylarının büyük bir kısmının kısa hikâyelerdeki sorulara vermiş oldukları cevaplarla bilimin doğasıyla ilgili boyutlarda naif veya orta düzey sayılabilecek bir anlayışa sahip olduğu söylenebilir.

Sonuç ve Tartışma

Öğretmen adaylarının bilimin doğası hakkındaki görüşlerini belirlemek için yapılan bu çalışmada kullanılan kısa hikâyeler yöntemi literatürde, Case (1997), Bodur (2003), Harlan (1996) ve Taşar'ın (2006) çalışmalarında bir ölçme değerlendirme aracı olarak kullanıldığı gibi, Roach (1993) ve Chan'ın (2000) çalışmalarında ise bir öğretim aracı olarak ta kullanılabilir. Ayrıca, Taşar'ın (2006) çalışmada görülmektedir. Çalışmanın sonucunda dikkatli seçilmiş bir kısa hikâyenin katılımcıların görüşlerini ölçmek için güçlü bir araç olabileceği vurgulanmaktadır. Ayrıca söz konusu bu çalışmada elde edilen sonuçlara bakılırsa katılımcıların

bilimin doğası hakkındaki görüşlerini herhangi bir yönlendirmeye ya da sınırlamaya maruz bırakmadan ortaya koymaya imkân sağladığı görülmektedir. Ancak katılımcıların görüşlerini analiz etmek ve rubrikler yoluyla sayısal veriler elde etmek bakımından hali hazırdaki bu çalışmanın alanda bir ilk olarak kullanıldığı söylenebilir. Çünkü bu çalışmada birden fazla bilimin doğası boyutu yine birden fazla kısa hikâyeye çalışılarak; oluşturulan rubriklerle katılımcıların görüşlerine sayısal bir değer verilmiştir. Böylece oluşturulan kısa hikâye ve rubrik ikilisiyle, katılımcıların bu konu hakkındaki görüşlerini ölçen likert tipi ya da çoktan seçmeli ölçeklerin getirdiği bazı sınırlamaların ve eksikliklerin (Lederman, Wade & Bell, 2000; Lederman & O'Malley, 1990; Aikenhead & Ryan, 1992) azaltılması düşünülmektedir.

Bilimin doğasına ilişkin görüş ve tutumları ölçmeye yönelik hazırlanan likert tipi ya da çoktan seçmeli ölçeklerin çoğu başarılı bulunmamaktadır (Lederman, Wade & Bell, 2000). Bu ölçeklerin dayandıkları bir bilim tanımı olduğu unutulmamalıdır. Bu yüzden de ölçeklerin ne anlamda bilimi kullandıkları ve bilim anlayışındaki çeşitlilik dikkate alınmalıdır. Nicel ölçeklerinin çoğunun geçerliliği yeterli görülmemektedir (Aikenhead & Ryan, 1992). Bu nedenle süreç içerisinde alanda daha çok nitel ağırlıklı ve açık uçlu soruların yer aldığı ölçme değerlendirme yöntemleri ortaya çıkmaya başlamıştır. Bunlardan sıklıkla kullanılanlara örnek olarak *Nature of Science Questionnaire* [NOS] ve *Views of Nature of Science Questionnaire* [VNOS] ölçme araçları sayılabilir.

Nature of Science Questionnaire [NOS] 1998 yılında Abd-El-Khalick tarafından bilimin doğasına ilişkin görüşleri ölçmek üzere geliştirilmiştir. Bu ölçek 8 açık uçlu sorudan oluşmuştur. Ölçeğin geçerlilik çalışması için yarı yapılandırılmış mülakatlar yapılmıştır. Ayrıca tüm mülakatlar iki farklı kişi tarafından birbirinden bağımsız bir şekilde değerlendirilmiş ve sonuçlar arasında % 87 arasında tutarlılık olduğu gözlenmiştir (Abd-El-Khalick, 2001). Hali hazırdaki bu çalışmada da elde edilen sonuçlar incelendiğinde farklı kısa hikâyelerde aynı ya da benzer konular hakkında verilen cevapların tutarlılığının yüksek olduğu görülecektir (tablo 5, 7, 9, 11, 13). Ayrıca NOS ölçeğinde yer alan konudan bağımsız altı sorunun varlığı düşünüldüğünde kısa hikâyelerin katılımcılara daha uygun uyarıcılar ve focus noktaları sağladığı (Jeffries ve Maeder, 2005) ve bu nedenle bu konular hakkındaki görüşlerini daha detaylı bir şekilde ortaya koymaya imkân sağladığı anlaşılmıştır.

Bilimin doğasına ilişkin görüşleri ölçmek üzere geliştirilmiş açık uçlu sorulardan oluşturulmuş bir diğer araç Lederman, Abd-El-Khalick, Bell, ve Schwartz (2002) tarafından geliştirilmiş ve mülakatlar ile desteklenen *Views of Nature of Science Questionnaire* [VNOS] ölçeğidir. Bu ölçek ortaöğretim ve üniversite öğrencileri, öğretmen adayları ve fen öğretmenlerinden oluşan bir gruba uygulanmıştır. Bunun yanı sıra mülakatlar yapılmıştır. Ölçeğin geçerliliği sürekli uzman görüşleri alınarak test edilmiş ve ölçek (VNOS –A, B, C) zaman içerisinde geliştirilmiştir. Her ne kadar Lederman ve diğ. (2002) tarafından geliştirilen VNOS anketi diğerlerine göre daha açık uçlu bir formatta hazırlanmış olsa da bu ankette de sorular içinde belli yönlendirmeler ve sınırlamalar görmek mümkündür (Taşar, 2006). Ayrıca bu anket sonrasında bir de görüşme yapmanın gerekliliği düşünülürse VNOS' un belli avantajlarının yanında dezavantajları olduğunu da söylemek mümkündür. Bununla birlikte VNOS anketinin düzey belirleyici değerlendirmeğe (summative assessment) imkân vermediğini de göz önünde bulundurmak gerekir.

Literatürde bilimin doğasına ilişkin görüşleri değerlendiren ölçek sayısı oldukça fazladır. Bu ölçeklerden bazıları kapalı bazıları ise açık uçlu olarak hazırlanmıştır. Açık uçlu ölçeklerin bireylerin düşünce zenginliğini istedikleri gibi ifade etmelerine olanak tanımaları bu tür ölçeklerin avantajını oluşturmaktadır. Ancak ölçmede standardın sağlanmasındaki güçlük ve sayının sınırlı olması birer dezavantajdır. Diğer yandan kapalı uçlu ölçekler ile daha fazla kişiye ulaşılabilir. Bu ölçeklerde ölçme daha standart bir şekilde yapılabilir. Ancak düşüncelerin istenildiği gibi ifade edilmesine olanak sağlanmaması, fikir zenginliğini sınırlamaktadır. Dolayısıyla hangi tür ölçeğin kullanılacağı araştırmacıya ve araştırmanın amacına bağlıdır. Her iki çeşidinde avantajları ve dezavantajları bulunmaktadır. Buna göre önemli olan ölçeklerin sınırlılıklarını bilmek ve araştırmanın amacına uygun olan ölçekleri seçebilmektir.

Son söz olarak bilim ve bilimin doğası gibi tanımları üzerinde herhangi bir konsensus sağlanamamış olan kavramlara ilişkin görüşleri ortaya koymak için açık uçlu ve cevaplayanları yönlendirmeyen ancak içerisinde barındırdığı olay, olgu, kişi ve zaman unsurlarıyla katılımcılara yararlı uyarıcılar sağlayan kısa hikâyelerin oldukça kullanışlı oldukları görülmüştür.

Kaynakça

- Abd-El-Khalick, F. & Lederman, N. G. (2000). Improving science teachers' conceptions of nature of science a critical review of the literature. *Journal of Science Education*, 22(7), 665-701.
- Abd-El-Khalick, F. (2001). Embedding nature of science instruction in preservice elementary science courses: Abandoning scientism, but...*Journal of Science Teacher Education*, 12(3), 215-233.
- Abd-El-Khalick, F., Waters, M., Le, A. P. (2008). Representations of Nature of Science in High School Chemistry Textbooks over the Past Four Decades. *Journal of Research In Science Teaching*. 45(7), 835-855.
- Aikenhead, G. S. (1988). An Analysis of Four Ways of a Assessing Student Beliefs About STS Topics. *Journal of Research in Science Teaching*. 25(8). 607- 629.
- Aikenhead, G. S., and Ryan, A. G. (1992). The development of a new instrument: "Views on science - technology - society"(VOSTS). *Science Education*, 76(5), 477- 491.
- American Association for the Advancement Of Science (AAAS). (1993). *Benchmarks for science literacy: A Project 2061 report*. New York: Oxford University Press.
- Bell, R. L. (2008). *Teaching the Nature of Science through Process Skills*. Boston: Allyn and Bacon.
- Bodur, Y. (2003). *Preservice teachers' learning of multiculturalism in a teacher education program*. Unpublished doctoral dissertation, The Florida State University, Tallahassee, FL.
- Büyüköztürk, Ş., Bökeoğlu, Ö. Ç., ve Köklü, N. (2009). *Sosyal bilimler için istatistik*. Ankara: Pegem Akademi Yayınları.
- Case, S.L. (1997). *Implementation of cooperative learning: Teacher beliefs and attitudes*. Unpublished doctoral dissertation, University of California - Santa Barbara, Santa Barbara, CA.
- Chan, K.S. (2000). *The impacts of infusing the interactive historical vignettes into 10th grade high school science instruction in Taiwan on Student understanding of the nature of science and science achievement*. Unpublished doctoral dissertation, The University of Texas at Austin, Austin, TX.
- Haidar, A. H. (1999). Emirates pre-service and in-science teachers' views about the nature of science. *International Journal of Science Education*, 21(8), 807-822.
- Harlan, S.L. (1996). *Exploring early childhood education students' beliefs about art education*. Unpublished doctoral dissertation, Rutgers, The State University of New Jersey, New Brunswick, NJ.
- Jeffries, C. & Maeder, D. W. (2005). Using Vignettes To Build and Assess Teacher Understanding of Instructional Strategies. *The Professional Educator*, 27(1- 2), 17- 28.
- Lederman, N. G., and O'Malley, M. (1990). Students' perceptions of tentativeness in science: Development, use, and sources of change. *Science Education*, 74(2), 225-239.
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29(4), 331-359.
- Lederman, N. G., Wade, P., and Bell, R. L. (2000). Assessing Understanding of the Nature of Science: A Historical Perspective. In W. F. McComas (Ed.), *The Nature of Science in Science Education* (pp 331- 350). Dordrecht, Boston, London: Kluwer Academic Publishers.
- Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., and Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward a valid and meaningful assessments of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, 39(6), 497- 521.
- Millî Eğitim Bakanlığı. (2005). *İlköğretim 6. ve 7. ve 8. Sınıf Fen ve Teknoloji Dersi Öğretim Programları*. Ankara: Millî Eğitim Bakanlığı Talim ve Terbiye Kurulu Başkanlığı.
- McComas, W. F., Clough, M. P. ve Almazroa, H. (2000). The role and the character of the nature of science. Bulunduğu eser: W. F. McComas (ed), *The Nature of Science in Science Education: Rationales and Strategies* (331-350). Dordrecht, Kluwer Academic Publishers.
- National Research Council. (1996). *National Science Education Standards*. Washington, D.C: National Academy Press.
- National Science Teachers Association. (1982). *Science- Technology- Society: Science education for the 1980s*. Washington, DC: Author.

- Özçelik, D. A. (1997). *Test Hazırlama Kılavuzu*. Genişletilmiş Üçüncü Baskı, ÖSYM Eğitim Yayınları, Ankara.
- Özgelen, S. (2010). *Exploring The Development Of Pre-Service Science Teachers' Views On Nature Of Science In Inquiry-Based Laboratory Instruction*. Unpublished doctoral dissertation, Middle East Technical University, Ankara.
- Roach, L.E. (1993). *Use of the history of science in a nonscience majors course: Does it affect students understanding of the nature of science?* Unpublished doctoral dissertation, Louisiana State University, Baton Rouge.
- Rubba, P. A., Andersen, H. O. (1978). Development of an instrument to assess secondary school students' understanding of the nature of scientific knowledge. *Science Education*, 62(4), 449-458.
- Taşar, M. F. (2006). Probing preservice teachers' understandings of scientific knowledge by using a vignette in conjunction with a paper and pencil test. *Eurasia Journal of Mathematics, Science and Technology Education*, 2(1), 53- 70.

Extended Abstract

The aim of this study is to determine the effectiveness of the short stories (vignette) for assessing the pre service science teachers' views about the nature of science. For this purpose, the validity and reliability studies of vignettes and rubrics used in the research are given.

The history of the assessment of the beliefs about the NOS reflects the evolution that has occurred in teacher thinking research. Traditionally, the dominant strategies employed with regard to the assessment of individuals' beliefs regarding the NOS have been those associated with quantitative methodology. Lederman, Wade and Bell (1998) report that the majority of researchers tended to develop instruments that allowed for easily "graded" and quantified measures of individuals' understandings. Some of these instruments involved some open-ended questioning; however, little emphasis was placed on providing an expanded view of an individual's beliefs regarding the NOS. Especially during the last 10 years, qualitative approaches have been more widely employed by researchers to assess NOS understandings. In this movement, interviews are considered to be crucial in yielding the essential data (Lederman, 1992) and, the use of such qualitative probes is considered to be important for generating profiles of the meanings individuals ascribe to various aspects of teaching and learning (Taşar, 2006).

Vignettes so far were utilized to collect data in various forms in studies from different areas in education in recent years. Their advantages particularly in qualitative, in-depth, small sample studies are obvious (Taşar, 2006). Case (1997) investigated the relationship between teacher beliefs and attitudes and the implementation of cooperative learning. She showed a videotaped vignette of teacher's classroom practice and interviewed the participants subsequently about its meaning for them. Bodur (2003) investigated preservice teachers' beliefs and attitudes about teaching culturally and linguistically diverse students. As part of his data collection strategy he used three multicultural vignettes in addition to other data sources.

This research comprised two stages: the pilot and the authentic application. The piloting was carried out in 2009 spring term with the participation of 90 pre service science teachers who had taken History and Nature of Science course while the main study included 79 pre service teachers in 2010 spring term. . In addition, consulted experts' opinions were used for validity and reliability.

In order to perform the validity analysis of the rubrics, the rubric drafts were sent to experts together with the participants' (n=45) first and tenth vignette responses. The experts were asked to compare the responses in the vignettes with the statements in the rubric. On the basis of experts' assessment and suggestions, necessary arrangements were made on the drafts. Although the reliability analysis of the rubrics was run in parallel with the validity analysis, here they will be explained as separate stages. The literature review suggested that the rubrics' reliability is usually calculated through inter-rater reliability (Bodur, 2003). In order to determine whether the two experts perceive the same situations from the rubric statements, the participants' responses to the first and tenth vignettes and the rubrics were sent together. The experts read participants' responses and scored them in terms of developmental, testability, and creativity. After this process was completed for the first and tenth vignettes responses, the researcher also scored the same responses. Thus, the reliability analysis was performed on the basis of three scores given by three different experts. For

the reliability analysis of the rubrics, calculating the Pearson's correlation between scorers' sets of scores was found to be appropriate. Consequently, the inter-rater reliability coefficient between the experts and between the experts and the researcher was found to be higher than .70. According to Bodur (2003) this result is satisfactory for the rubrics' reliability. On the basis of this, it could be argued that the rubrics are perceived by everybody in the same way, and the scores obtained by means of these rubrics are independent from the scorers. In brief, it could be argued that the rubrics used in the study are reliable.

On the basis of the rubrics which were found to be reliable and valid, participants' opinions on each of the dimensions in each vignette were obtained and scored. In order to obtain data on the reliability of the vignettes used for study purposes, the relationships and consistent within the quantitative data were examined. To this end, the rubric scores were analyzed by means of Wilcoxon Signed Ranks Test. To identify pre-service teachers' opinions on the nature of science, the rubric scores which determine the assessment of responses to the vignettes are highly important. Considering the developmental dimension, on the basis of findings from the first, second, and third vignettes (Table 4 and 5), and the sixth, seventh and eighth vignettes (Table 6 and 7), it could be argued that vignettes are reliable in terms of assessing and evaluating participants' opinions on this dimension. The same is also true for both the testability (Table 8; 9, 10, 11) and creativity dimensions (Table 12 and 13).

In order to establish the content validity of the vignettes used in the present study, five experts who were academicians were consulted. The formation process of the vignettes is explained in detail in the methodology section. The vignettes which were used in the study based on expert opinion were found to have content validity. It was concluded that both the issues explained in vignettes and the related questions cover concepts and principles related to the nature of science. The Pearson's correlation coefficient between the participants' total scores from the rubrics and the term final scores was found to be 0.86. The calculated coefficient was found to be higher than 0.70. This finding shows that the vignettes are satisfactory in terms of predictive validity.

In conclusion, vignettes, which are open-ended and do not direct the participants in terms of giving opinions on concepts like nature and nature of science on which there is a lack of consensus, provide useful stimuli for the participants with the events, phenomena and time elements they incorporate.