

PAPER DETAILS

TITLE: Fen Bilimleri Öğretmenlerinin Üst Düzey Öğrenme Düzeyini Ölçen Soru Gelistirme

Öz-Yeterlik Ölçegi: Geçerlik ve Güvenirlik

AUTHORS: Ahmet BOLAT,Özgen KORKMAZ,Sevilay KARAMUSTAFAOGLU

PAGES: 372-416

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/1312667>


<http://kefad.ahievran.edu.tr>

Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi

ISSN: 2147 - 1037

Question Development Self-Efficacy Scale Measuring High Level Learning Level of Science Teachers: Validity and Reliability

Ahmet Bolat
Özgen Korkmaz
Sevilay Karamustafaoğlu

Article Information



CrossMark

DOI:10.29299/kefad.800889

Received: 27.09.2020

Revised: 11.01.2021

Accepted: 18.02.2021

Keywords:

Validity and Reliability,
Science Teacher,
Scale Development,
Self-efficacy,
High Level Learning.

Abstract

The aim of the study is to develop a question development self-efficacy scale that measures science teachers' high-level learning. The population of the research science teachers in secondary schools are in the Ministry of National Education in Turkey. The sample of the study consists of 518 science teachers who voluntarily participated among these teachers. While creating the item pool of the scale, the literature was first reviewed. A trial form of the scale consisting of 53 items was prepared. The item pool of the scale was presented to experts from different fields. After the necessary corrections, the scale trial form was applied to science teachers. At the end of the application, exploratory factor analysis, item-total correlation analysis and independent samples t test were used to determine the construct validity of the scale. According to the results of factor analysis; 23 items were removed from the scale form. According to the result of the exploratory factor analysis, it was determined that the scale consists of a single factor. By conducting a confirmatory factor analysis, it was determined that the model formed was verified by the data. Cronbach's alpha reliability value of the scale was calculated as 0.977. In addition, the stability of the scale was calculated by performing a test-retest application. Considering the values made and calculated, it was concluded that the scale is valid and reliable.

Fen Bilimleri Öğretmenlerinin Üst Düzey Öğrenme Düzeyini Ölçen Soru Geliştirme Öz-Yeterlik Ölçeği: Geçerlik ve Güvenirlik

Makale Bilgileri



CrossMark

DOI: 10.29299/kefad.800889

Yükleme: 27.09.2020

Düzeltilme: 11.01.2021

Kabul: 18.02.2021

Anahtar Kelimeler:

Geçerlik ve Güvenirlik,
Fen Bilimleri Öğretmeni,
Ölçek Geliştirme,
Öz-yeterlik,
Üst Düzey Öğrenme.

Öz

Çalışmanın amacı, fen bilimleri öğretmenlerinin üst düzey öğrenme düzeyini ölçen soru geliştirme öz-yeterlik ölçeği geliştirmektir. Araştırmanın evreni Türkiye'deki Millî Eğitim Bakanlığı bünyesindeki ortaokullarda görevli fen bilimleri öğretmenleridir. Çalışmanın örneklemini bu öğretmenler arasından gönüllü olarak katılan 518 fen bilimleri öğretmeni oluşturmaktadır. Ölçeğin madde havuzu oluşturulurken ilk olarak alanyazın taraması yapılmıştır. 53 maddeden meydana gelen ölçek deneme formu hazırlanmıştır. Ölçeğin madde havuzu, farklı alanlardan uzmanların görüşüne sunulmuştur. Gerekli düzeltmelerden sonra ölçek deneme formu fen bilimleri öğretmenlerine uygulanmıştır. Uygulama sonunda ölçeğin yapı geçerliğini belirlemek için açımlayıcı faktör analizi, madde-toplam korelasyon analizi ve bağımsız örneklem için t testi yapılmıştır. Faktör analizleri sonucuna göre; 23 madde ölçek formundan çıkarılmıştır. Açımlayıcı faktör analizi sonucuna göre ölçeğin tek faktörden meydana geldiği belirlenmiştir. Doğrulayıcı faktör analizi yapılarak oluşan modelin veriler tarafından doğrulandığı tespit edilmiştir. Ölçeğin Cronbach alfa güvenirlik değeri 0,977 olarak hesaplanmıştır. Ayrıca ölçeğin kararlılığı test-tekrar test uygulaması yapılarak hesaplanmıştır. Yapılan ve hesaplanan değerler göz önünde bulundurulduğunda ölçeğin geçerli ve güvenilir olduğu sonucuna ulaşılmıştır.

Sorumlu Yazar: Ahmet BOLAT, Öğretmen, İl Millî Eğitim Müdürlüğü Ölçme Değerlendirme Merkezi, Türkiye, ahmbolat@yahoo.com, ORCID ID: 0000-0002-3581-2899

Özgen KORKMAZ, Prof. Dr., Amasya Üniversitesi, Türkiye, ozgenkorkmaz@gmail.com, ORCID ID: 0000-0003-4359-5692

Sevilay KARAMUSTAFAOĞLU, Prof. Dr., Amasya Üniversitesi, Türkiye, sevilayt2000@yahoo.com, ORCID ID: 0000-0002-2852-7061

Atıf için: Bolat, A., Korkmaz, Ö. & Karamustafaoğlu, S. (2021). Fen bilimleri öğretmenlerinin üst düzey öğrenme düzeyini ölçen Soru Geliştirme Öz-Yeterlik Ölçeği: Geçerlik ve güvenirlik. *Kırşehir Eğitim Fakültesi Dergisi*, 22(1), 372- 416.

Giriş

Günümüzdeki eğitim yaklaşımları bilgileri araştırmadan, sorgulamadan alıp sadece tüketen bireyler yerine; araştıran, sorgulayan, bilgiye nasıl ulaşacağını bilen, bilgiye ulaşma ilkelerini uygulayarak bilgiyi üreten ve bunu hem kendi yaşantısına tatbik edip hem de insanlığın kullanımına sunabilen bireyler yetiştirmeyi amaçlamaktadır (Akpınar ve Aydın, 2010). Buna bağlı olarak yeni eğitim yaklaşımlarında öğrencilerin bilgileri ezberlemesi yerine bilgileri öğrenmeleri ve bu bilgileri içselleştirmesi gerekmektedir. Bu durum öğrencilerin üst düzey öğrenmelerini gerektirmektedir (Doğanay, 1997). Bloom Taksonomisi'ne göre bilişsel alanda gerçekleştirilen öğrenme altı basamakta gerçekleşir (Anderson ve diğerleri, 2001). Bilgi, kavrama, uygulama alt düzey, analiz, sentez ve değerlendirme üst düzey olarak ifade edilir (Şahinel, 2002). Üst düzey öğrenme ve düşünme birbirini tamamlamaktadır. Alt düzeydeki öğrenme, önceki öğrenmeler ve akılda kalan bilgileri, formülleri tekrarlama olarak ifade edilirken, üst düzey öğrenme, öğrencileri analiz etmeye, yorum yapmaya ve bilgiyi harekete geçirme yeteneklerini geliştirmeye yönlendirir (Anderson ve Krathwohl, 2010). Öğrencileri üst düzey öğrenmeye yöneltmek için hazır bilgi vermek yerine, onların bilgiye ulaşma yollarına yöneltmek gerekir. Bu anlamda öğrencilerin gerçek yaşam görevleri ile öğrenmelerini sağlamakta fayda vardır.

Alt ve üst düzey öğrenme en yaygın haliyle yenilenmiş Bloom Taksonomisi'ne göre belirlenir. Bu taksonomide alt düzey öğrenmeler, hatırlama, anlama ve uygulama basamaklarından; üst düzey öğrenmeler ise çözümleme, değerlendirme ve oluşturma basamaklarından oluşmaktadır (Anderson ve diğerleri, 2001; Lewis ve Smith, 1993; Moon, 2004; Şahinel, 2002). Düşünme becerileri öğrenme becerilerinin bir göstergesidir. Birey hangi düşünme basamağındaysa aynı zamanda o öğrenme basamağındadır (Güneş, 2012). Çözümleme basamağı, bütünsel bir bilgiyi ya da bir şemayı, meydana getiren parçaları, yine o bütün ve şemada bulunduğu şekilde parçalarına ayırma işlemidir. Bununla birlikte ortaya atılan fikirler arası tutarlık ve geçerlik bağlantılarının da araştırılması bu basamağın kapsamındadır. Değerlendirme basamağı, tüm durumları dikkate alarak bir yargıya varma süreci olarak tarif edilebilir. Oluşturma basamağı, özellikleri belirli ilişki ve kurallar göz önünde bulundurularak bir bütün meydana getirme olarak tarif edilebilir. Fakat öğrencinin oluşturduğu tüm ürünler, bu kavrama karşılık gelmez (Bloom, Engelhart, Furst, Hill ve Krathwohl, 1956; Bloom, Krathwohl ve Masia, 1984; Krathwohl, Bloom ve Masia, 1964). Oluşturmada yenilik, özgünlük, buluş, yaratıcılık gibi durumlar vardır (Simpson, 1966; Sosniak, 1994; Sönmez, 1993). Bu düzeydeki öğrenmelerin gerçekleşip gerçekleşmediğini belirlemek için ölçme araçlarında kullanılan soru maddelerinin çözümleme, değerlendirme ve oluşturma basamağında olması önem taşımaktadır. Ölçme araçlarını hazırlayan öğretmenlerin üst düzey öğrenmelerin ölçüleceği soru maddelerini hazırlama konusunda öz-yeterliğe sahip olması gerekmektedir. Türkiye'de hali hazırda doğrudan üst düzey öğrenmeleri ölçen soru yazımına yönelik bir ders yada eğitim programı bulunmamaktadır. Öğretmen adayları lisans eğitimleri boyunca sadece eğitimde ölçme ve değerlendirme dersi almaktadır. Bu ders

programının içeriğinde ise doğrudan üst düzey öğrenmeleri ölçen soru yazımına yönelik bir kazanım yer almamaktadır(YÖK,2021). Öğretmenlerin hizmetiçi eğitim programlarında üst düzey bilişsel becerilerin ölçülmesi ve soru hazırlama kursu yer almaktadır. Bu kursu alabilmek için açık uçlu soru hazırlama 1-2 şeklinde iki kursu almak gerekmektedir(MEB,2021). Öğretmenlerin hizmet içi kursları yaygın kitleye ulaşması ve üç ayrı kursu alma zorunluluğundan dolayı bu kursa öğretmenlerin erişimi oldukça zordur. Öğretmenlerin üst düzey öğrenmeleri ölçmek için hazırlayacağı sorularla ilgili öz-yeterliğe sahip olup olmadığının belirlenmesi gerekir. Bu da ancak geçerlik ve güvenirlik çalışması yapılmış bir psikometrik ölçme aracı olan ölçeklerle sağlanabilir. Alanyazın incelendiğinde ise böyle bir ölçeğin olmadığı görülmektedir.

Yukarıda açıklanan üst düzey öğrenmelerin, ne kadar gerçekleştiğinin belirlenmesi gerekmektedir. Üst düzey öğrenmelerin ne kadar gerçekleştiğini belirlemenin en etkili yollarından biri de üst düzey öğrenmeleri ölçen soruların geliştirilmesidir. Soru maddelerinde de sorunun bilişsel düzeyinin belirlendiği en yaygın haliyle Bloom Taksonomisi'dir (Anderson ve diğerleri, 2001; Anderson ve Krathwohl, 2010; Farr, 2010). Çözümleme basamağındaki bir soru maddesinde gerekli ve gereksiz verileri ayırt etmeyi veya herhangi bir konu ile ilgili veriler arasında hiyerarşi oluşturma gibi konunun ana hatlarını belirlemeyi ölçmek isteyen sorular yer almaktadır (Bloom, ve diğerleri, 1956; Lewis ve Smith,1993; Sosniak,1994). Değerlendirme basamağındaki soru maddeleri bir problemin çözümünde yapılan yanlışları denetlemeyi, problemin çözümünde birden fazla çözüm yollarından hangisinin daha doğru olacağını tespit etmeyi gerektiren kazanımları ölçmeyi amaçlayan soru maddeleridir (Bloom ve diğerleri, 1956; Moon, 2004; Sosniak,1994). Oluşturma düzeyinde yer alan soru maddeleri, daha önceden bilinmeyen bir çözüm yöntemi geliştirilmesi gereken ya da çözümü için yöntemler olmakla beraber bu yöntemlere alternatif yöntemler oluşturabilme becerisini ölçen soru maddeleridir (Ardahanlı, 2018; Bloom ve diğerleri, 1956; Farr, 2010; Sosniak, 1994). Çözümleme düzeyindeki sorularda öğrencinin sınıflama, betimleme, grupta, karşılaştırma, ayırt etme, ilişkilendirme, açıklama gibi becerileri yoklanır. Değerlendirme düzeyindeki sorularda öğrencinin akıl yürütme, önerme, birleştirme, planlama, formüle etme, sonuç çıkarma, sentezleme gibi beceriler yoklanır. Oluşturma düzeyindeki sorularda ise öğrencinin karar verme, yargılama, tercih etme, ispat etme gibi becerileri yoklanır (Anderson ve diğerleri, 2001; Anderson ve Krathwohl, 2010; Koray, Altunçekiç, ve Yaman, 2005).

Türkiye'de Dünya'daki gelişmelere paralel olarak öğretim programları sürekli yenilenmektedir. 2018 yılında yayınlanan son fen bilimleri programında üst bilişsel becerilerin öğrenilmesine ve bunlara yönelik ölçme ve değerlendirme faaliyetlerine vurgu yapılmaktadır (Millî Eğitim Bakanlığı [MEB], 2018). Eğitimde ölçme değerlendirme, öğrencilerin kazanımları edinme düzeylerini belirleyerek; eğitim süreçlerinin kontrol edilmesinde ve iyileştirilmesinde eğitimcilere geri dönüt vermektedir. Ölçme ve değerlendirme işlemleri, ölçme araçlarında yer alan sorular sayesinde gerçekleştirilmektedir (Turgut ve Baykul, 2015). Ölçme ve değerlendirme yaklaşımı da eğitim

yaklaşımlarının değişimine bağlı olarak değişmektedir. Bu da soruların niteliklerinin değişmesine neden olmaktadır. Son yıllara kadar sınavlarda kullanılan sorular öğrencilerin bilgilerini hatırlama ve tanımasını ölçerken; son yıllarda hazırlanan uygulanan sınavlardaki sorular öğrencilerin var olan bilgilerini çözümleme, değerlendirme ve oluşturma düzeylerini ölçmeye yöneliktir (Doğan, 2019). Gerek merkezi sınavlarda gerekse okullarda yapılan ölçme ve değerlendirme uygulamalarında soruları öğretmenler hazırlamaktadır. Bu noktada öğretmenlerin soru yazma yeterliliklerinin ortaya konulması eğitimde ölçme ve değerlendirme faaliyetlerinin sağlıklı yürütülmesi açısından önem kazanmaktadır.

Alanyazın incelendiğinde Türkiye’deki fen bilimleri öğretmen adaylarının hatırlama, anlama ve uygulama düzeyindeki öğrenmeleri ölçen soru yazmada yeterli iken; üst düzey öğrenmeyi gerektiren çözümleme, değerlendirme ve oluşturma düzeyini ölçen soru yazmada yetersiz olduğu anlaşılmaktadır (Koray, Altunçekiç, ve Yaman, 2005; Özcan ve Akcan, 2010). Benzer durum aktif görevde olan fen bilimleri öğretmenleri için de geçerlidir. Görevde olan fen bilimleri öğretmenlerinin de hazırladıkları sorular daha çok alt düzey öğrenme düzeyini ölçen sorulardır (Akpınar ve Ergin, 2006; Mutlu ve diğerleri, 2003; Cansüğü Koray ve Yaman, 2002; Dindar ve Demir, 2006; Ayvaci ve Şahin, 2009). Fen bilimleri öğretmen adayları ve öğretmenlerin hazırladıkları soruların düzeyini belirleyen çalışmalar doküman incelenmesi sonucu belirlenmiştir. Belirleme işleminde geçerlik ve güvenirlik çalışması yapılan ya da uyarlanmış bir psikometrik araç olan ölçek kullanılmamıştır. Doküman incelemeleri doğası gereği geçerlik ve güvenirlik açısından tarama desenlerine göre daha zayıftır (Cansız Aktaş, 2019:127). İşte bu noktada öğretmenlerin üst düzey öğrenmeleri ölçen soru hazırlama öz-yeterlik düzeylerinin belirlenmesinde geçerlik ve güvenirlik çalışması yapılmış ölçeklere ihtiyaç duyulmaktadır.

Öz-yeterlik, Bandura (1977) tarafından geliştirilen sosyo bilişsel kuramla ortaya çıkmıştır. Bandura (1977) öz-yeterlik kavramını kişinin bir işi belli seviyede başarmasına ilişkin inancı şeklinde tanımlamıştır. Diğer taraftan öz-yeterlik, herhangi bir işteki becerileri ortaya koymada gerekli olan eylemin yapılması ya da organize edilmesi için kişinin kendi yeteneklerine olan inancıdır (Bandura, 1997). Bunun yanında öğretmen öz yeterliği, öğretim amaç ve hedefleri için etkili yöntem ve teknikleri uygulama kapasitesine yönelik öğretmenin inanışıdır (Park ve Oliver 2008). Öz-yeterlik kuramına göre, yüksek öz-yeterlik durumu, öğrencilerinin gelişimi için uğraşan daha kararlı öğretmenlerin göstergesidir (Woolfolk, Winne, Perry ve Shapka, 2009). Öz yeterliliğin iki temel bileşeni vardır. Bunlar daha önceki yaşam deneyimlerine dayalı bir eylemin ürünleri ve karşılaştığı sorunları çözmede kendi yeteneklerine olan inancıdır (Bandura, 1997).

Doğrudan fen bilimleri öğretmenlerine yönelik Türkiye’de, STEM görüş ölçeği (Gül Biçer, Uzoğlu ve Bozdoğan, 2018), ortaöğretime geçiş sistemi ile ilgili “fen bilimleri öğretmeni görüş ölçeği” (Taşkın ve Aksoy, 2018), fen bilimleri öğretmenlerinin laboratuvar ve öğretim sürecine ilişkin kaygı ölçeği (Kahraman ve Polat, 2017), öğretim programına göre görüş ölçeği (Temli Durmuş ve Ok, 2014), ilköğretim fen ve matematik öğretmenleri için ölçme ve değerlendirme yeterlilik ölçeği (Yıldırım Ekinci

ve Köksal, 2011), kişisel mesleki gelişim ölçeği (Bilgin ve Balbağ, 2016) geliştirildiği anlaşılmıştır. Diğer taraftan Dibise ve Mcdonald (2015) geliştirdiği araştırma ve sorgulamaya dayalı öğretim hakkında fen bilimleri öğretmenlerinin tutumları ölçeğinin (Açıkgöz, Uluçınar Sağır ve Ozan, 2018) tarafından uyarlandığı anlaşılmıştır.

Araştırmanın Amacı

Alanyazın incelemesi sonucu doğrudan fen bilimleri öğretmenlerine yönelik bir ölçek geliştirme çalışmasının sınırlı sayıda yapıldığı anlaşılmıştır (Biçer, Uzoğlu ve Bozdoğan,2018; Durmuş Bakioğlu,2013; Taşkın ve Aksoy,2018; Çevik,2017). Diğer taraftan sadece fen bilimleri öğretmenlerinin ölçme ve değerlendirme ya da soru yazmalarına yönelik geliştirilmiş bir ölçeğe alanyazında rastlanmamıştır. Bu çalışma ile fen bilimleri öğretmenlerine yönelik ölçme ve değerlendirme öz-yeterlikleri ile ilgili alanyazına ölçek kazandırılacağı düşünülmektedir. Ölçekler psikometrik ölçme araçlarından olması nedeniyle geçerliği ve güvenirliği çalışma grubunun özelliklerinden etkilenmektedir (Büyüköztürk,2012). Öğretmen adayları ile görevde olan öğretmenlerin öz-yeterlik algılarındaki farklılaşmanın çok fazla olacağı düşüncesinden dolayı ölçeğin sadece fen bilim öğretmenlerine yönelik olması araştırmacılar tarafından uygun görülmüştür. Bu sayede fen bilimleri öğretmenlerine yönelik ölçek geliştirme çalışmalarının sayısal ve çeşitlilik bakımından zenginleştirileceği düşünülmektedir. Bu amaçla fen bilimleri öğretmenlerinin üst düzey soru geliştirme öz-yeterliklerini belirlemek için ölçek geliştirme çalışması yapılmıştır.

Yöntem

Bu çalışma nicel kökenli, betimsel tarama modeli olarak tasarlanmış bir ölçek geliştirme çalışmasıdır. Betimsel tarama modeli geçmişteki veya günümüzde hali hazırdaki bir durumu aynı şekilde açıklamayı hedefleyen bir araştırma yöntemidir (Karasar, 2000). Bu çalışmada fen bilimleri öğretmenlerinin üst düzey öğrenme düzeyini ölçen soru geliştirme öz-yeterliklerini belirleyen ölçek geliştirilmek istendiğinden çalışmanın modeli tarama modelidir.

Örneklem/Çalışma Grubu/Katılımcılar

Araştırmanın evreni Türkiye'deki Millî Eğitim Bakanlığı'na bağlı ortaokullarda görev yapan fen bilimleri öğretmenleridir. Çalışmanın örneklemi bu öğretmenler arasından gönüllü olarak katılan 518 fen bilimleri öğretmeni oluşturmaktadır. 1. uygulamada 265, 2.uygulamada 227, Test Tekrar Test uygulanasında 26 olmak üzere toplam 518 fen bilimleri öğretmenleri oluşturmaktadır. Örneklemde yer alan öğretmenlerin bazı demografik özelliklerinin dağılımı Tablo 1'de sunulmuştur.

Tablo 1. Çalışma grubunu oluşturan öğretmenlerin demografik özelliklerinin betimsel dağılımı

Özellikler		1.Uygulama		2.Uygulama		Test Tekrar Uygulaması		Genel Toplam	
		f	Y(%)	f	Y(%)	f	Y(%)	f	Y(%)
Cinsiyet	Erkek	121	45,66	103	45,37	16	61,50	240	46,33
	Kadın	144	54,34	124	54,63	10	38,50	278	53,67
	Toplam	265	100,00	227	100,00	26	100,00	518	100,00
Kıdem	0-5 Yıl	52	19,62	38	16,74	0	0,00	90	17,37
	6-10 Yıl	82	30,94	76	33,48	6	23,10	164	31,66
	11-15 Yıl	59	22,26	53	23,35	9	34,60	121	23,36
	16-20 Yıl	31	11,70	22	9,69	4	15,40	57	11,00
	21-25 Yıl	25	9,43	17	7,49	5	19,20	47	9,07
	26 Yıl ve üzeri	16	6,05	21	9,25	2	7,70	39	7,53
	Toplam	265	100,00	227	100,00	26	100,00	518	100,00
Öğrenim Düzeyi	Ön lisans	1	0,38	0	0,00	1	3,80	2	0,39
	Lisans	198	74,72	181	79,74	18	69,20	397	76,64
	Yüksek lisans	61	23,01	39	17,18	6	23,10	106	20,46
	Doktora	5	1,89	7	3,08	1	3,80	13	2,51
	Toplam	265	100,00	227	100,00	26	100,00	518	100,00
Görev Yapılan İl	Çorum	26	9,81	25	11,01	18	69,20	69	13,32
	İstanbul	22	8,30	27	11,89	0	0,00	49	9,46
	Sivas	36	13,59	5	2,21	0	0,00	41	7,92
	Amasya	21	7,92	13	5,73	3	11,50	37	7,14
	Diğer	160	60,38	157	69,16	5	19,30	322	62,16
	Toplam (68 Farklı İl)	265	100,00	227	100,00	26	100,00	518	100,00

f:Frekans, Y:Yüzde

Tablo 1 incelendiğinde; öğretmenlerin cinsiyet bakımından kadın öğretmenlerin daha fazla olmakla birlikte oransal anlamda çok fazla farklılık olmadığı anlaşılmaktadır. Öğretmenlerin kıdemi açısından çalışma grubu değerlendirildiğinde; en yüksek oranın 6-10 yıl arası kıdemi olan öğretmenlerin (%31,66), en düşük oranın ise 26 yıl ve daha fazla kıdemi olan öğretmenlerin (%7,53) olduğu anlaşılmaktadır. Öğrenim düzeyi açısından ise büyük çoğunluğunun lisans mezunu (%76,64) olduğu anlaşılmaktadır. Çalışma grubunun görev yaptığı şehirler açısından değerlendirildiğinde ise en fazla katılımın sırasıyla Çorum, İstanbul, Sivas ve Amasya illerinden olduğu anlaşılmaktadır. Bununla birlikte çalışma grubunda toplam 68 farklı ilden fen bilimleri öğretmeni yer almaktadır.

Veri Toplama Araçları/Veri Toplama Yöntemleri / Madde Havuzunun Oluşturulma Süreci

Ölçeğin geliştirilmesi aşamasında ilk olarak alanyazın taraması yapılmıştır. Bu çerçevede fen bilimleri öğretmenlerinin ölçme değerlendirme durumlarını araştıran çalışmalar incelenmiştir. Bu inceleme sonucu fen bilimleri öğretmenlerinin üst düzey soru geliştirme öz-yeterliğinin ortaya konulması ihtiyacı olduğu belirlenmiştir. Daha sonra öz-yeterlik kavramı ile ilgili çalışmalar incelenmiştir. Böylece öz-yeterlik kavramının alanyazın alt yapısı belirlenmiştir (Bandura,1977; Park ve Oliver 2008; Woolfolk ve diğerleri, 2009). Üst düzey öğrenme ve düşünme kavramları ve göstergelerinin nasıl oluşturulacağı alanyazın taraması ile belirlenmiştir (Anderson ve diğerleri, 2001). Elde edilen bilgilerden yola çıkılarak üst düzey öğrenmenin Bloom taksonomisinde yer alan bilişsel süreç

basamaklarından çözümleme, değerlendirme ve oluşturma basamaklarından oluşması ve bu nedenle araştırmacılar tarafından üç basamaktaki öğrenme kazanımlarını ölçme öz-yeterliliğine yönelik madde havuzu oluşturma çalışması başlamıştır. İlk aşamada 46 maddeden oluşan madde havuzu oluşturulmuştur. Madde havuzunda çözümleme basamağına yönelik 19, değerlendirme basamağına yönelik 15 ve yaratma basamağına yönelik 12 madde yer almıştır. Maddeler oluşturulurken Bloom Taksonomisi'ndeki çözümleme, değerlendirme ve oluşturma basamaklarında yer alan soruların özellikleri araştırılmıştır. Araştırmalar sonucunda bu üç basamakta yer alan soruların öğrencilerde hangi beceri ve kazanımları yokladığı belirlenmiştir. Bu beceri ve kazanımlara ait sınıflama, birleştirme, yorumlama, ilişkilendirme, önerme, değişkenleri belirleme, ölçme vb. anahtar kelimeler ölçek maddeleri için üst düzey öğrenmenin göstergesi olarak belirlenmiştir. Bu çerçevede madde havuzu oluşturulmuştur. Oluşturulan madde havuzu, fen bilimleri öğretmenlerinin, maddelerde belirtilen öz-yeterlik düzeylerini belirlemek için beşli likert tipinde düzenlenmiştir. Bu seçenekler; kesinlikle katılmıyorum(1), katılmıyorum(2), kararsızım(3), katılıyorum (4) ve kesinlikle katılıyorum (5) olarak hazırlanmış ve derecelendirilmiştir. İki fen eğitimi uzmanı, iki dil uzmanı, bir eğitim bilimleri uzmanı ve iki ölçme-değerlendirme uzmanının görüşüne sunularak görünüş ve kapsam geçerliliği sağlanmaya çalışılmıştır. Uzman görüşü sonucunda 3 maddenin benzer özelliği ölçen maddeler olduğu gerekçesiyle ölçekten çıkarılmıştır. Bir fen eğitimi uzmanının önerisi sonucunda ölçeğe 10 madde eklenmiştir. Dil ve ölçme-değerlendirme uzmanlarının görüşleri doğrultusunda gerekli değişiklikler yapılmış ve 53 maddelik ölçek pilot uygulamaya hazır hale getirilmiştir. Ölçeğin pilot uygulaması, 5 fen bilimleri öğretmeni ile yapılmıştır. Pilot uygulamadaki öğretmenler örneklemini yansıtan grup olmasına dikkat edildi ve pilot uygulamaların sonuçları esas analize dahil edilmemiştir. Öğretmenlerin uyarıları doğrultusunda içerik aynı kalmakla birlikte dil ve anlam bilgisi bakımından düzenleme yapılarak esas uygulamaya hazır hale getirilmiştir.

Verilerin Analizi

İstatistiki analizler kapsamında ölçek için elde edilen verilere dayanarak, ölçeğin yapı geçerliğini incelemek üzere ilk olarak verilerin faktör analizine uygun olup olmadığını belirlemek için, KMO ve Bartlett test analizleri yapılmıştır. Analiz sonucunda verilerin faktör analizine uygun olması nedeniyle veriler üstünde açımlayıcı faktör analizi (AFA) ile doğrulayıcı faktör analizi (DFA) yapılmış; ölçek maddelerinin faktörlere dağılımı temel bileşenler analizi ile tespit edilmiş; Varimax Dik Döndürme tekniği ile de faktör yükleri belirlenmiştir. AFA, geliştirilecek ölçme aracının yapısı hakkında bilgi edinilmesini, aynı yapıyı ölçen değişkenleri olabildiğince az sayıda faktör ile açıklamayı sağlayan istatistiksel bir analiz türüdür (Balcı, 2013). Temel Bileşenler Analizi neticesinde, faktör yük değerleri 0,30'dan küçük olan maddelerin ve iki ya da daha fazla faktörden de yük alıp, aldığı yük değerleri arasında 0,10 ve daha az fark olan maddelerin ölçekten çıkarılması gerekmektedir (Büyüköztürk, 2012). Bununla birlikte, faktör yük değeri 0,40'ın altında olup ölçekten çıkarılması ölçeğin güvenilirliğini önemli oranda etkilemiyorsa; bu maddelerin de çıkarılması ölçeği daha geçerli ve

güvenilir yapılabilir. Ölçek maddelerinin faktör yük değerlerinin 0,30'dan büyük olması, genel varyansın ise tek faktörlü ölçeklerde en az %30, çok faktörlü ölçeklerde ise en az %40'ını açıklaması, davranış bilimleri açısından yeterlidir (Kline, 1994; Scherer, Wiebe, Luther ve Adams, 1988). Faktör analizi sonuçları değerlendirilirken esas ölçüt faktör yükleridir (Gorsuch, 1983; Balcı, 2013). Faktör yük değerlerinin yüksek olması, değişkenin ilgili faktör altında bulunabileceğinin göstergesidir (Büyüköztürk, 2012). Ayrıca çok faktörlü ölçekler için ortak faktör varyansının hesaplanması önemlidir ve temel bileşenler analizi ile faktörlerin her bir değişken üzerinde neden oldukları ortak varyans olarak ifade edilmektedir (Çokluk, Şekercioğlu ve Büyüköztürk, 2010). Ortak faktör varyansının ise 0,50'den büyük olması beklenir (Erkuş, 2016). Bu bağlamda çalışmada ortak faktör varyans değeri 0,50'den, faktör yükü 0,40'dan küçük olan ve birden fazla faktöre dağılıp faktör yükleri arasında 0,10 ve daha az fark olan maddeler uzman görüşlerine de başvurularak ölçekten çıkarılmıştır.

Açımlayıcı faktör analizi sonucunda ortaya çıkan ölçek formu ilk uygulamanın çalışma grubundan farklı bir çalışma grubuna uygulanıp, elde edilen veriler veriler üzerinde doğrulayıcı faktör analizi yapılmıştır. Gözlenen ile örtük değişkenler arasındaki ilişkilerin birer hipotez olarak değerlendirilip test edilmesi ilkesine dayanan analiz yöntemi Doğrulayıcı Faktör Analizidir (Pohlmann, 2004). Doğrulayıcı faktör analizinde en çok olabilirlik (maximum likelihood) tekniği kullanılmıştır. En çok olabilirlik (maximum likelihood) tekniği hem esnek hem de anlaşılabilirliğinin kolay olması nedeniyle tercih edilmiştir. Doğrulayıcı faktör analizi ile oluşan ölçek modelinde gözlenen değerlerin Tablo 2'de belirtilen aralıklarda olması beklenir (Kline, 2005; Sümer, 2000; Şimsek, 2007).

Tablo 2. Doğrulayıcı faktör analizi ile oluşan ölçek modelinde gözlenen değerlerin uyum indisleri aralıkları

Uyum İndisleri	Mükemmel Uyum	Kabul Edilebilir Uyum
χ^2/d	$\chi^2/d < 3$	$\chi^2/d < 5$
RMSEA	$0 < RMSEA < 0.05$	$0.06 \leq RMSEA < 0.08$
S- RMR	$0 \leq S- RMR \leq 0.05$	$0.06 \leq S- RMR \leq 0.08$
NNFI	$0.97 \leq NNFI \leq 1$	$0.90 \leq NNFI \leq 0.96$
CFI	$0.97 \leq CFI \leq 1$	$0.90 \leq CFI \leq 0.96$
GFI	$0.95 \leq GFI \leq 1$	$0.90 \leq GFI \leq 0.96$
AGFI	$0.95 \leq AGFI \leq 1$	$0.90 \leq AGFI \leq 0.96$
IFI	$0.95 \leq IFI \leq 1,00$	$0.90 \leq IFI \leq 0.95$

Faktör analizi ile elde edilen 30 maddelik ölçeğin madde ayırt ediciliği bağımsız örneklemeler için t testi ile analiz edilmiştir. Ölçeğin tüm maddelerinin ayrı ayrı ölçeğin genel amacına hizmet etme düzeyleri madde-toplam korelasyonları Pearson's r testi ile analiz edilmiştir. Tek faktörlü ölçeklerde tüm maddelerden ayrı ayrı hesaplanan puan ile ölçeğin genelinden hesaplanan puan arasında korelasyonun pozitif ve anlamlı düzeyde olması, ölçeğin maddelerinin tamamın ayrı ayrı, faktörün dolayısıyla da ölçeğin genel amacına hizmet etme düzeyinin belirlenmesi yönüyle bir ölçüt olarak kullanılmaktadır (Balcı, 2013). Ölçeğin güvenilirliğini tespit etmek üzere, iç tutarlılık katsayıları ile kararlılık katsayıları hesaplanmıştır. İç tutarlılık düzeyinin hesaplanmasında Cronbach alpha güvenilirlik katsayısı, Guttman split-half, iki eş yarılar korelasyon katsayısı ve Sperman-Brown

katsayılarından yararlanılmıştır. Güvenirlik katsayısının en az 0,70 olması, ölçeğin güvenilir olduğunun bir göstergesidir (Gorsuch, 1983; Büyüköztürk, 2002). Ölçeğin kararlılık düzeyi ise 30 gün arayla yapılan iki uygulamanın sonuçları arasındaki korelasyonu hesaplanarak iki uygulama arasındaki ilişki belirlenmiştir. Bir ölçeğin kararlı olabilmesi için belirli aralıklarla yapılan iki uygulamanın neticeleri arasında kabul edilebilir düzeyde tutarlılık olması gerekir. Tutarlılık düzeyini ifade eden güvenirlilik katsayısı 0 ile 1 arasında değişir ve 1'e yaklaştıkça tutarlılık düzeyi yükselir (Gorsuch, 1983). Korelasyon katsayıları için $r \leq 0,30$ düşük düzeyde, $r \leq 0,70$ orta düzeyde, $r \leq 1,00$ düzeyi ise yüksek korelasyonu göstermektedir (Büyüköztürk, 2002).

Etik İle İlgili Hususlar

Yapılan bu çalışmada “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir. Çalışma öncesinde aşağıda bilgileri verilen etik kurulu onayı alınmıştır.

Etik kurul izin bilgileri:

Etik değerlendirmeyi yapan kurul adı	: Amasya Üniversitesi Sosyal Bilimler Etik Kurulu
Etik değerlendirme kararının tarihi	: 15.05.2020
Etik değerlendirme belgesi sayı numarası	: 30640013-044

Bulgular

Çalışmanın bu bölümünde ölçeğin geçerlik ve güvenirlilik analizleri kapsamında yapılan işlemler ve bulgulara yer verilmiştir.

Ölçeğin Geçerliğine Dair Bulgular

Çalışmanın bu bölümünde “fen bilimleri öğretmenlerinin üst düzey öğrenme düzeyini ölçen soru geliştirme öz-yeterlik ölçeği”nin geçerliğini belirlemek amacıyla yapı geçerliği, madde ayırt ediciliği ve madde-toplam korelasyonları analiz edilmiş ve bulgular sunulmuştur:

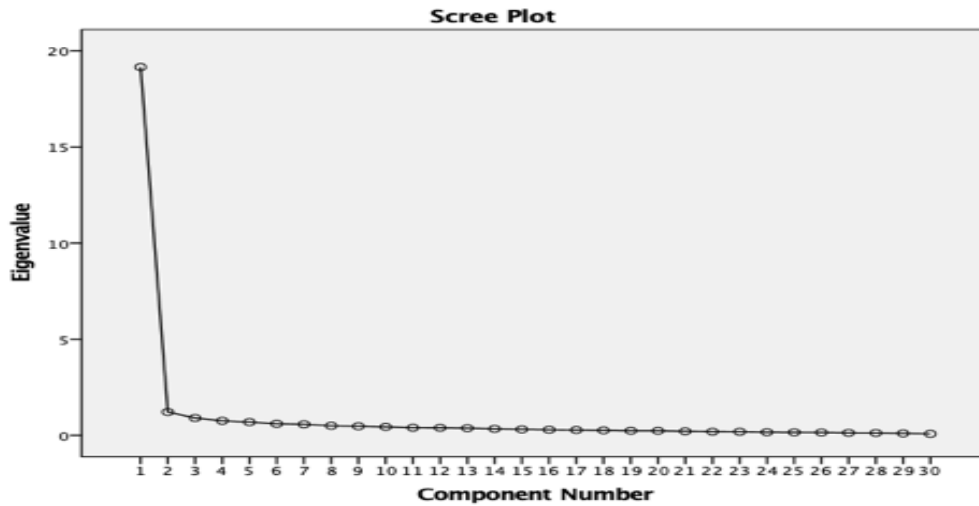
Yapı geçerliği

Açımlayıcı Faktör Analizine Dair Bulgular: Ölçeğin yapı geçerliğini belirlemek için veriler, öncelikle Kaiser-Meyer-Okin (KMO) ve Bartlett test analizlerine tabi tutulmuş ve KMO= 0,966; Bartlett testi değeri ise $\chi^2 = 14150,785$; $sd=1378$ ($p=0,000$) olarak tespit edilmiştir. Bu sonuçlar kapsamında, 53 maddelik ölçeğin verilerinin faktör analizine uygun olduğu anlaşılmıştır.

Ölçeğin faktör analizine uygun olduğu anlaşılması üzerine, ölçeğin kaç boyutlu olduğunu belirlemek için temel bileşenler analizi yapılmıştır. Temel bileşenler analizi, faktörleştirme tekniklerden en çok kullanılan tekniklerden birisidir (Büyüköztürk, 2002). Sonrasında temel bileşenleri açısından Varimax Dik Döndürme Tekniği kullanılmıştır. Bu doğrultuda faktör yükü 0,40'ın altında olan 1 ve faktör yükü 0,1 daha az farkla birden fazla faktörlere yayılan 10 olmak üzere toplam 11 madde ölçekten

çıkarılarak geriye kalan 42 madde üzerinden tekrar faktör analizi yapılmıştır. İkinci kez yapılan analiz sonucunda ölçekte kalan 42 maddenin, bir faktör altında öbeklendiği anlaşılmıştır. Bunun üzerine 42 madde üzerinden üçüncü kez bir faktörlü ölçek yapısı olacak şekilde yeniden açımlayıcı faktör analizi yapılmıştır. Bu analizler sonucunda ortak varyans değeri 0,50'den küçük olan 8 madde ve faktör yükü 0,40'tan küçük olan 4 madde daha ölçekten atılmıştır. Böylece; ölçekten toplam 23 madde atılmış ve son durumda 30 maddeli bir faktörlü ölçek elde edilmiştir. Atılan maddelerin kapsam geçerliliğini bozmamasını sağlamak için oluşan yeni madde havuzu tekrar iki fen eğitimi uzmanına incelenmiştir. Bu iki uzman ilk değerlendirmeyi yapan uzmanlardır. Fen eğitimi uzmanlarının, dahil edilmeyen 23 maddenin atılmasının kapsam geçerliliğini olumsuz etkilemediği doğrultusundaki görüşlerinin ardından diğer analizlere geçilmiştir. Son durumda 30 maddeli ölçeğin KMO değeri 0,967; Bartlett Testi değeri $\chi^2=8057,031$; $sd=435$; $p=0,00$ olarak hesaplanmıştır. Ölçekte geriye kalan 30 maddenin rotasyona tabi tutulmaksızın) faktör yüklerinin 0,713 ile 0,856 arasında olduğu belirlenmiştir. Diğer yandan ölçek kapsamında yer alan maddelerin ve tek faktörün toplam varyansın %63,861'ünü açıkladığı tespit edilmiştir. Alanyazın incelendiğinde faktör yüklerinin 0,30'tan küçük olamaması ve sosyal bilimlerde açıklanan varyans yüzdesinin tek faktörlü yapılarda %30 ve üzerinde, çok faktörlü yapılarda ise %40 ve üzerinde olması yeterli kabul edilmektedir (Büyüköztürk, 2012; Eroğlu, 2008). Ölçeğin yapısı itibarıyla 30 maddelik tek faktör, "fen bilimleri öğretmenlerinin üst düzey soru geliştirme öz-yeterlik ölçeği" şeklinde isimlendirilmiştir.

Yukarıda açıklanan durum, faktörlerin öz değerlerine göre çizilen yamaç birikinti grafiğinde Grafik 1'de sunulmuştur.



Grafik 1. Yamaç birikinti grafiği (faktörlere göre özdeğerler)

Grafik 1'de, birinci faktörden sonra yüksek ivmeli düşüşün olduğu; bu sebeple varyansa bir faktörün yüksek oranda katkı sunduğu; bununla birlikte diğer faktörlerin birinci faktöre göre varyansa çok az katkı sunduğu, çizgilerin birinci kırılma noktasından sonra yatay görüntü almaya başladığı, diğer bir deyiş ile varyansa olan katkılarının birbirine yakın olduğu anlaşılmaktadır (Büyüköztürk, 2012; Eroğlu, 2008).

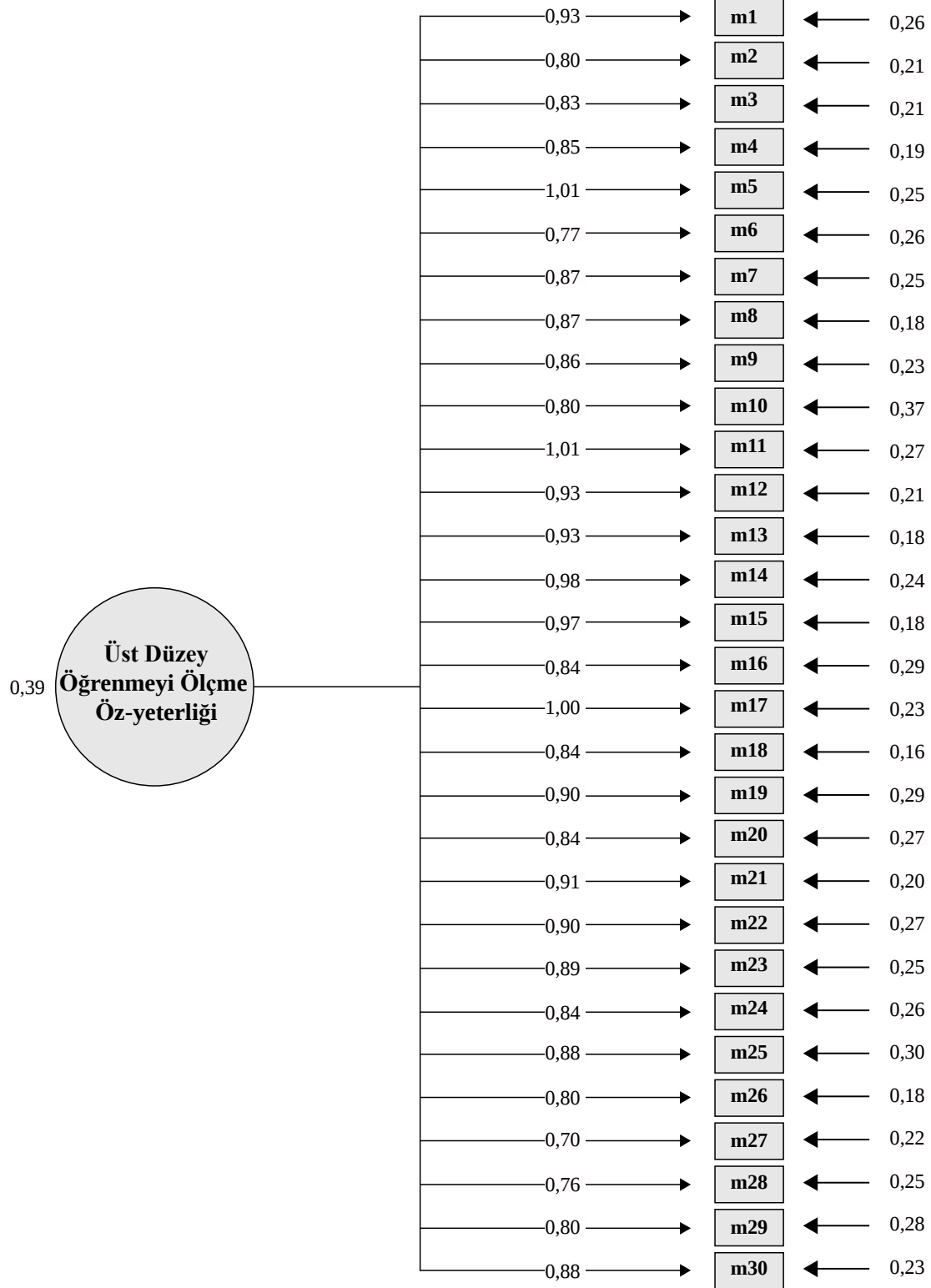
Yapılan işlemler neticesinde, ölçekte kalan 30 maddenin tek faktör bakımından madde yükleri ile faktörün özdeğeri ve varyansı açıklama oranına dair bulgular Tablo 3.'de belirtilmiştir.

Tablo 3. Ölçeğin tek faktöre göre yapılan faktör analizi sonuçları

Faktör	Madde No	Ortak Varyanslar	Ortak Faktör Varyansı	Faktör Yüklü	Madde No	Ortak Varyanslar	Ortak Faktör Varyansı	Faktör Yüklü
Üst Düzey Öğrenmeyi Ölçme Özyeterliliği	M1	0,585	0,585	0,856	M16	0,689	0,689	0,809
	M2	0,678	0,678	0,848	M17	0,650	0,650	0,806
	M3	0,666	0,666	0,845	M18	0,698	0,698	0,797
	M4	0,732	0,732	0,842	M19	0,532	0,532	0,786
	M5	0,662	0,662	0,840	M20	0,657	0,657	0,783
	M6	0,606	0,606	0,837	M21	0,714	0,714	0,778
	M7	0,613	0,613	0,835	M22	0,584	0,584	0,771
	M8	0,701	0,701	0,833	M23	0,594	0,594	0,765
	M9	0,655	0,655	0,830	M24	0,575	0,575	0,764
	M10	0,635	0,635	0,828	M25	0,509	0,509	0,758
	M11	0,709	0,709	0,825	M26	0,686	0,686	0,745
	M12	0,719	0,719	0,823	M27	0,617	0,617	0,739
	M13	0,705	0,705	0,816	M28	0,547	0,547	0,730
	M14	0,681	0,681	0,814	M29	0,511	0,511	0,715
	M15	0,694	0,694	0,810	M30	0,556	0,556	0,713
Özdeğer								19,158
Açıklanan Varyans								63,861

Tablo 3 incelendiğinde ölçeğin 30 madde ve tek faktör içerdiği; maddelerin faktör yüklerinin 0,713 ile 0,856 arasında değiştiği görülmektedir. Bununla birlikte tek faktörün ölçekteki öz değeri 19,158; varyansa katkı oranının ise %63,861 olduğu görülmektedir. Ayrıca tüm maddelerin ortak faktör varyans değerinin 0,50'den yüksek olduğu görülmektedir.

Doğrulamalı Faktör Analizine Dair Bulgular: Açımlayıcı faktör analizinin ardından tek faktörden meydana geldiği belirlenen ölçeğin faktör yapılarının doğrulanabilmesi için farklı bir örneklem olan 227 fen bilimleri öğretmeninden toplanan veriler üstünde doğrulamalı faktör analizi yapılmıştır. Herhangi bir sınırlamaya gidilmeksizin en fazla olabilirlik (maximum likelihood) tekniği kullanılarak gerçekleştirilen doğrulamalı faktör analizi sonucunda elde edilen model Şekil 1'de sunulmuştur.



Şekil 1. Ölçeğin doğrulayıcı faktör analizi bağlantı diyagramı

Şekil 1’de çizilen yol diyagramında, her bir maddenin faktörü, dolayısıyla ölçeği hangi oranda temsil ettiğini belirten standardize edilmiş değerler görülmektedir. Model incelendiğinde maddelerin hata varyansının 0,18 ile 0,30 arasında değiştiği; gizil değişken ile gözlenen değişken arasındaki yapısal katsayının ise 0,70 ile 1,01 arasında değiştiği anlaşılmaktadır.

Doğrulayıcı faktör analizi sonucu hesaplanan uyum iyiliği indisi değerleri Tablo 4’te sunulmuştur.

Tablo 4. Doğrulamalı faktör analizi sonucu elde edilen uyum indisi değerleri

Uyum İndisleri	İndis Değerleri
χ^2/d	2,657
RMSEA	0,84
S- RMR	0,025
NNFI	0,828
CFI	0,884
GFI	0,768
AGFI	0,729
IFI	0,885

Tablo 4 incelendiğinde uyum indisi değerlerinden $\chi^2/d=2,657$, RMSEA=0,840, S-RMR=0,025, NNFI= 0,828, CFI=0,884, GFI=0,768, AGFI=0,729 ve IFI=0,885 olarak hesaplanmıştır. Uyum iyiliği indisleri değerlendirildiğinde modelin χ^2/d iyi uyum, RMSEA kabul edilebilir uyum ve S-RMR değeri ise mükemmel uyum aralıklarında olduğu söylenebilir. NNFI, CFI, GFI, AGFI ve IFI uyum iyiliği indisleri ise kabul edilebilir değer aralıklarında değildir. Bununla birlikte en az üç uyum iyiliği indis değerlerinin modelin veriler tarafından doğrulanması için yeterlidir (McDonald ve Ho, 2002). Bu sonuçlar da modelin veriler tarafından doğrulandığını göstermektedir.

Madde Ayırt Ediciliği: Ölçekteki maddelerin ve ölçeğin tamamının ayırt ediciliği incelenmiştir. Bu kapsamda ölçeğin tamamından hesaplanan puanların %27'lik alt ve üst grupları arasında anlamlı farklılık olup olmadığı incelenmiştir. Bu amaca yönelik her bir maddenin puanları ve toplam puanlar üzerinden bağımsız örneklem için t testi yapılmıştır. Analiz sonuçları Tablo 5.'te sunulmuştur.

Tablo 5. Ölçeğin madde ayırt edicilik özelliklerine yönelik bağımsız örneklem t testi sonuçları

Madde Numarası	Ortalama		t	Madde Numarası	Ortalama		t
	Alt Grup	Üst Grup			Alt Grup	Üst Grup	
m1	3,25	4,54	-14,924	m16	3,39	4,63	-14,383
m2	3,52	4,74	-14,436	m17	3,10	4,56	-17,557
m3	3,48	4,71	-14,538	m18	3,47	4,66	-15,366
m4	3,44	4,68	-15,033	m19	3,18	4,44	-13,478
m5	3,13	4,57	-16,947	m20	3,35	4,57	-13,691
m6	3,65	4,80	-12,549	m21	3,28	4,56	-15,212
m7	3,32	4,56	-14,316	m22	3,06	4,39	-14,534
m8	3,32	4,59	-14,716	m23	3,29	4,53	-14,606
m9	3,58	4,80	-14,158	m24	3,21	4,44	-14,171
m10	3,65	4,79	-11,953	m25	3,12	4,38	-14,069
m11	3,20	4,66	-16,616	m26	3,38	4,50	-14,238
m12	3,21	4,59	-16,481	m27	3,35	4,45	-13,663
m13	3,33	4,60	-15,161	m28	3,45	4,47	-12,076
m14	3,23	4,458	-16,393	m29	3,49	4,61	-13,729
m15	3,33	4,61	-15,108	m30	3,11	4,47	-17,223
Toplam					105,85	146,338	-22,76

N=492; **=p<0,05

Tablo 5 incelendiğinde tüm maddelerin ve ölçeğin genelinin alt ve üst grup puanları arasında anlamlı farklılık (p<0,05) olduğu anlaşılmaktadır. Bu durum ölçeğin genelinin ve maddelerinin tamamının ayrı ayrı ayırt edici özelliğe sahip olduğunu göstermektedir.

Ayrıca madde-toplam korelasyonu tekniğine göre maddelerinin tamamının ayrı ayrı, ölçeğin genelinden ulaşılan puanlar arasındaki korelasyonlar hesaplanarak tüm maddelerin ayrı ayrı genel amaca hizmet etme düzeyi belirlenmiştir. Her bir madde için hesaplanan madde-faktör korelasyon değerleri Tablo 6’da sunulmuştur.

Tablo 6. Madde-faktör (toplam) puanları korelasyon analizi

Madde Numarası	r	Madde Numarası	r
m1	0,761*	m16	0,778*
m2	0,788*	m17	0,802*
m3	0,789*	m18	0,818*
m4	0,818*	m19	0,734*
m5	0,799*	m20	0,775*
m6	0,743*	m21	0,818*
m7	0,769*	m22	0,757*
m8	0,816*	m23	0,765*
m9	0,782*	m24	0,749*
m10	0,734*	m25	0,723*
m11	0,820*	m26	0,801*
m12	0,823*	m27	0,740*
m13	0,821*	m28	0,720*
m14	0,809*	m29	0,701*
m15	0,826*	m30	0,753*

N=492; *=p<0,001

Tablo 6 incelendiğinde maddelerin korelasyon değerlerinin 0,701 ile 0,826 arasında değiştiği, tüm ilişkilerin pozitif yönlü ve anlamlı olduğu anlaşılmaktadır. Bu durum her bir maddenin geçerli ve ölçeğin geneli ile tutarlı olduğunu, dolayısıyla ölçeğin genel amacına hizmet ettiğini göstermektedir.

Ölçeğin Güvenirliğine Dair Bulgular

Ölçeğin güvenilirliğini belirlemek için veriler üstünde iç tutarlılık ve kararlılık analizleri yapılmıştır. Bu işlemler ile elde edilen bulgular aşağıda gösterilmiştir:

İç tutarlılık düzeyi: Ölçeğin güvenilirlik analizi; eş iki yarı arasındaki korelasyon değeri, Cronbach alpha güvenilirlik katsayısı, Guttman split-half ve Sperman-Brown formülü yoluyla hesaplanmıştır. Ölçeğin güvenilirliğine yönelik yapılan bu testlerden elde edilen bulgular Tablo 7’de sunulmuştur.

Tablo 7. Ölçeğin güvenilirlik analizi sonuçları

Faktör(Ölçek)	N	Madde Sayısı	İki Eş Yarı Korelasyonları	Sperman Brown	Guttman Split-Half	Cronbach Alpha
Üst Düzey Öğrenmeyi Ölçme Özyeterliği	492	32	0,902	0,948	0,948	0,977

Tablo 7 incelendiğinde tek faktör ve 30 maddeden oluşan ölçeğin eş iki yarı korelasyonları 0,902; Sperman Brown güvenilirlik katsayısı 0,948; Guttman Split-Half değeri 0,948; Cronbach alpha güvenilirlik katsayısı ise 0,977 olarak hesaplanmıştır.

Kararlılık düzeyi: Güvenilir bir ölçeğin araçlarının özelliklerinden birisi kararlılıktır. Kararlı bir ölçeğin aracı belirli zaman örneğin, en az üç hafta arayla aynı örneklem grubuna uygulandığında iki uygulama arasında pozitif yönlü ve anlamlı bir ilişki olması gerektiği vurgulanmaktadır (Balci, 2013). Ölçeğin

kararlılık seviyesi, test tekrar test yöntemi yoluyla belirlenmiştir. Ölçeğin 30 maddelik nihai hali, uygulamanın yapıldığı 26 öğretmene 30 gün arayla iki kez uygulanmıştır. İki uygulama sonucunda hesaplanan puanlar arasındaki tutarlılık, korelasyon yoluyla hem tüm maddelerin ayrı ayrı hem de ölçeğin geneli bakımından incelenmiştir. Bu sayede ölçekte yer alan her maddenin ayrı ayrı ve ölçeğin tamamının kararlılığı test edilmiştir. Elde edilen bulgular Tablo 8.'de sunulmuştur.

Tablo 8. Ölçeğin maddelerinin test-tekrar test sonuçları

Madde Numarası	r	Madde Numarası	r
m1	0,641**	m16	0,772**
m2	0,687**	m17	0,545**
m3	0,627**	m18	0,770**
m4	0,725**	m19	0,587**
m5	0,393*	m20	0,725**
m6	0,735**	m21	0,657**
m7	0,656**	m22	0,547**
m8	0,449*	m23	0,556**
m9	0,548**	m24	0,715**
m10	0,836**	m25	0,700**
m11	0,478**	m26	0,489**
m12	0,637*	m27	0,589**
m13	0,806**	m28	0,700**
m14	0,38	m29	0,773**
m15	0,643**	m30	0,613**
Toplam		0,833*	

N: 26; *= $p<0,05$; **= $p<0,01$

Tablo 8 incelendiğinde ölçeği oluşturan maddelerin tamamı, test-tekrar test yöntemi yoluyla hesaplanan korelasyon katsayılarının 0,380 ile 0,836 arasında değişim gösterdiği ve her bir maddeye ait ilişkinin 14. madde dışında anlamlı ve pozitif olduğu görülmüştür ($p<0,01$ veya $p<0,05$). 14. maddenin korelasyonu pozitif olmakla birlikte anlamlı düzeyde bir ilişki belirlenememiştir. Ancak uzman görüşleri doğrultusunda bu maddenin kapsam geçerliliği açısından önemli olduğu yönündeki dönütleri nedeniyle maddenin ölçekte kalmasına karar verilmiştir. Ölçeğin tamamına ait kararlılık düzeyine yönelik elde edilen bulguda ölçeğin tamamına ait korelasyon katsayısı pozitif ve anlamlıdır ($r=0,833$ ve $p<0,05$). Güvenilir bir ölçme aracı; kararlı, tutarlı ve duyarlı olmalıdır. Bu anlamda kararlılık katsayısı biçiminde hesaplanan değerler, ölçeğin güvenilir olduğunun bir göstergesi şeklinde değerlendirilebilir (Hovardaoğlu, 2000). Bu açıdan geliştirilen ölçek değerlendirildiğinde, istenilen özellikte bir ölçek olduğu söylenebilir.

Sonuçlar, Tartışma ve Öneriler

Fen bilimleri öğretmenlerinin üst düzey öğrenmeyi ölçen soru geliştirme öz-yeterlik ölçeği geliştirmek amacıyla gerçekleştirilen bu çalışmada ilgili ölçek; geçerli ve güvenilirliği yüksek, likert tipinde, tek faktörlü ve 30 maddeden oluşmaktadır. Maddelerin tümü; Kesinlikle Katılmıyorum (1), Katılmıyorum (2), Kararsızım (3), Katılıyorum (4), Kesinlikle Katılıyorum (5) olarak derecelendirilmiştir. 30 maddeli ölçeğin KMO değeri 0,967; Bartlett Testi değerleri $\chi^2=8057,031$; $sd=435$; $p=0,00$ 'dır. Ölçeğin 30 maddesinin rotasyona tabi tutulmaksızın (unrotated) faktör yükleri 0,713 ile 0,856

arasındadır. Ölçekte yer alan maddelerin ve tek faktörün toplam varyansın %63,861'ünü açıklamaktadır. Ayrıca tüm maddelerin ortak faktör varyans değeri 0,50'den büyüktür. Ölçeğin doğrulayıcı faktör analizi sonucunda hesaplanan uyum indisi değerlerinden $\chi^2/d=2,657$, RMSEA=0,840, S-RMR=0,025, NNFI=0,828, CFI=0,884, GFI=0,768, AGFI=0,729 ve IFI=0,885 şeklindedir. Uyum iyiliği indisleri değerlendirildiğinde ölçeğe ait modelin veriler tarafından doğrulanmış model olduğu anlaşılmaktadır. Çünkü modelin veriler tarafından doğrulanabilmesi için çok sayıda uyum indisi türleri vardır. Bu indislerden en üç tanesinin kabul edilebilir düzeyde olması yeterli görülmektedir (McDonald ve Ho,2002). Ölçeğin madde analizi sonucunda tüm maddelerin ve ölçeğin genelinin alt ve üst grup puanları arasında anlamlı farklılık ($p<0,05$) vardır. Bu nedenle ölçeğin geneli ve tüm maddeler ayırt edici özelliğe sahiptir (Balci, 2013; Erkuş, 2016). Ölçeğin madde-toplam korelasyon analizi sonucuna göre maddelerin korelasyon değerleri 0,701 ile 0,826 arasında değişmekte, tüm ilişkiler pozitif yönlü ve anlamlıdır. Bu nedenle her bir maddenin geçerli ve ölçeğin geneli ile tutarlıdır, dolayısıyla her bir madde ölçeğin genel amacına hizmet etmektedir. Ölçeğin eş iki yarı korelasyonları 0,902; Sperman Brown güvenirlik katsayısı 0,948; Guttman Split-Half değeri 0,948; Cronbach Alpha güvenirlik katsayısı ise 0,977'dir. Bu nedenle ölçeğin güvenirliği çok yüksektir (Büyüköztürk, 2002). Ölçeğin tüm maddelerinin test-tekrar test yöntemi yoluyla hesaplanan korelasyon katsayıları 0,380 ile 0,836 arasındadır ve her bir maddeye ait ilişki 14. madde dışında anlamlı ve pozitifdir ($p<0,01$ veya $p<0,05$). Ancak uzman görüşleri doğrultusunda 14. madde kapsam geçerliliği bakımından önemli olduğu için ölçekte yer almaktadır. Ölçeğin tamamına ait kararlılık düzeyine ilişkin korelasyon katsayısı pozitif ve anlamlıdır ($r=0,833$ ve $p<0,05$). Ölçeğin psikometrik özellikleri beraber değerlendirildiğinde ölçeğin geçerli ve güvenilir olduğu sonucuna ulaşılabilir.

Alanyazın incelendiğinde son on yılda ortaokul 4-8. sınıf fen ve matematik dersi veren öğretmenlerin ölçme değerlendirme yeterlilikleri ölçeği (Yıldırım Ekinci ve Köksal, 2011) ve öğretmenlerin-ölçme değerlendirme yeterlilikleri ölçeği (Kuzu, 2016) gibi ölçeklere rastlanmakla birlikte, öğretmenlerin soru yazma becerilerini konu alan bir ölçeğe rastlanılmamıştır. Bu çalışmada geliştirilen ölçeğin fen bilimleri öğretmenlerinin üst düzey öğrenmeyi ölçen soru geliştirme öz-yeterliliklerinin belirlenmesi açısından önemli olduğu sonucuna varılmıştır. Ölçeğin geçerlik özelliği üç farklı analiz ile incelenmiştir. Bunlar, faktör analizi, madde-toplam korelasyon analizi ve %27'lik alt grup ile üst grubu ölçeğin ayırt etme özelliklerini analiz eden bağımsız örneklem için t testi analizleridir. Ölçek geliştirme çalışmaları incelendiğinde faktör analitik yöntemlerden en yaygın kullanılan teknik temel bileşenler analizidir. Alanyazında yer alan ölçek geliştirme çalışmalarında yapı geçerliliği sadece açımlayıcı faktör analizi yoluyla yapılmaktadır (Erkuş, 2016). Bu çalışmada da alanyazınla uyumlu bir şekilde açımlayıcı faktör analizi kapsamında yapılan temel bileşenler analizi ile; ölçeğin tek faktörden oluştuğu anlaşılmaktadır. Her bir maddenin faktör yükleri, faktörün öz değeri ve açıklanan varyans oranı incelendiğinde ölçeğin, yapı geçerliği olan bir ölçek olduğu ifade edilebilir.

Alanyazın incelendiğinde ölçek geliştirme çalışmalarının büyük çoğunluğu, yapı geçerliliğini sadece açımlayıcı faktör analizi yoluyla incelemektedir (Gül ve Sözbilir, 2015). Gül ve Sözbilir'e (2015) göre bu durumun nedeni; AFA'nın SPSS gibi yaygın bir şekilde bilinen ve kullanılan yazılımla yapılabilirken; DFA'nın LISREL ve Amos gibi az bilinen ve yaygın olmayan programlarda yapılabilmesi, bu nedenle bilgi, tecrübe eksikliğinden kaynaklanmaktadır. AFA geliştirilen psikometrik ölçüm aracının yapısını anlamaya, DFA ise tanımlanan bu yapıyı sınamaya yöneliktir. Her iki analizde ölçek geliştirme aşamasında önemlidir ve birbirini tamamlamaya analiz türüdür. Bu nedenle ölçek geliştirme aşamasında önce AFA ile ölçeğin yapısı belirlenmeli devamında DFA ile tanımlanan yapı test edilmelidir (Erkuş, 2016). Bu çalışmada ise açımlayıcı faktör analizi ile tek faktörden meydana geldiği belirlenen ölçeğin faktör yapısının doğrulanıp doğrulanmadığını belirlemek için doğrulayıcı faktör analizi yapılmıştır. Doğrulayıcı faktör analizi ile hesaplanan uyum indislerinden iyi, mükemmel ve kabul edilebilir olmak üzere üç değer bulunmaktadır. Uyum indislerinden en az üçünün kabul edilebilir ya da daha iyi düzeyde olması verilerin doğrulandığı anlamına gelir (McDonald ve Ho, 2002). Bu durumda yapılan doğrulayıcı faktör analizi sonuçları ile ölçek modelinin gözlenen değerinin, verilerle uyumlu olduğu, başka bir ifade ile oluşan bu modelin veriler tarafından doğrulandığını göstermektedir.

Ölçekte bulunan tüm maddelerin, ölçek ile ölçülmeye çalışılan özellikleri hangi düzeyde ölçeğildiğini tespit etmek için veriler üzerinde madde-toplam korelasyonları hesaplanmıştır. Hesaplanan değerlere göre her bir maddenin toplam puan ile korelasyonunun pozitif yönlü ve anlamlı düzeyde olduğu; ölçekte bulunan her bir maddenin, ölçeğin tamamı ile ölçülmek istenen özelliği ölçekleme hedefine hizmet ettiği sonucuna ulaşılmıştır. Ölçek geliştirme aşamasında madde analizi de çok önemlidir ve mutlaka maddelerin ayırt ediciliği incelenmelidir (Erkuş, 2016). Bu nedenle bu çalışmada maddelerin ayırt edici olup olmadığı sınanmıştır. Bağımsız örneklem için t testi ile ölçeğin her bir maddesinin ve genelinin ayırt edici özelliği olduğu sonucuna ulaşılmıştır. Alanyazında hazırlanan ölçek geliştirme çalışmalarında büyük çoğunlukla madde-test korelasyonu ve madde analizi yapılmaktadır (Şahin ve Boztunç Öztürk, 2018; Gül ve Sözbilir, 2015). Ölçeğin ölçüt geçerliliğini belirlemek için, alanyazın taraması gerçekleştirilmiş; içerik ve hedef açısından benzer bir ölçek bulunamadığı için ölçüt geçerliliği belirlenememiştir.

Ölçek geliştirme çalışmalarında güvenilirlik hesaplamalarında en sık tercih edilen analiz Cronbach α testidir (Gül ve Sözbilir, 2015; Tosun ve Taşkesenligil, 2015). Bu çalışmada ölçeğin iç tutarlılık katsayıları; Cronbach Alpha, Sperman-Brown formülü, iki eş yarı korelasyonları ve Guttman split-half güvenilirlik formülü ile hesaplanmıştır. Bu hesaplamalar sonucundaki değerler göz önünde bulundurulduğunda, ölçeğin güvenilirliğinin çok yüksek olduğu ve güvenilir ölçümler yapabildiği sonucuna ulaşılabilir.

Ölçeğin kararlılığını belirlemek için, 30 gün arayla yapılan uygulamalarla elde edilen veriler üzerinde test-tekrar test yöntemi ile iki uygulama arasındaki korelasyon analizi yapılmıştır. Bu analizler

tüm maddeler için ve ölçeğin tamamı için hesaplanmıştır. Hesaplanan korelasyon katsayılarına göre ölçekte bulunan maddelerden düşük düzeyde korelasyona sahip madde bulunmamaktadır. Maddelerden 21'i orta, 9'u ise yüksek düzeyde korelasyona sahiptir. Ayrıca ölçeğin genel korelasyonu da yüksek düzeydedir. 14. madde dışındaki tüm maddeler ve ölçeğin geneli pozitif yönlü (14. madde dahil), anlamlı düzeyde korelasyona sahiptir. Bu durum ölçeğin kararlı ölçüm yapabildiği sonucuna ulaşılmıştır. Türkiye'de hazırlanan ölçek geliştirme çalışmalarında çok az sayıda kararlılık analizi yapılmaktadır (Şahin ve Boztunç Öztürk, 2018).

Ölçekte yer alan maddeler değerlendirildiğinde, maddeler Bloom Taksonomisi'ndeki çözümleme, değerlendirme ve oluşturma basamaklarına yönelik öğretmenlerin soru geliştirebilme öz-yeterliğini ölçmektedir. Ölçek madde havuzu oluşturulurken Bloom Taksonomisi'ndeki bu basamaklar dikkate alınmıştır. Çünkü yaygın kabulle üst düzey öğrenme düzeylerini bu basamaklar temsil etmektedir (Anderson ve diğerleri, 2001; Anderson ve Krathwohl, 2010; Farr, 2010). Ölçek maddeleri Bloom Taksonomisi'nde yer alan çözümleme, değerlendirme ve oluşturma basamaklarında yer alan becerilerin göstergeleri dikkate alınarak uzman görüşleri doğrultusunda hazırlanmıştır. Çözümleme basamağında yer alan maddeler sınıflandırma, destekleme, delil toplama, yorumlama, sonuç çıkarma, ilişki kurma, parçalara ayırma, çıkarım yapma, veri toplama, verileri kaydetme, verileri analiz etme, yorumlama becerilerini yoklamaktadır (Koray, Altunçekiç ve Yaman, 2005). Değerlendirme basamağında yer alan maddeler hazırlanırken karşılaştırma, kanıtlama, tahmin etme, eleştirme, ölçme, kestirme, karar verme becerilerini yoklanmaktadır. Oluşturma basamağındaki maddeler ise ispatlama, önerme, düzenleme, organize etme, birleştirme, geliştirme, tasarlama, ispatlama, sunma, paylaşma becerilerini yoklamaktadır. Üst düzey öğrenmeye sahip öğrencilerin bu becerilere sahip olması gerekmektedir (Anderson ve diğerleri, 2001; Anderson ve Krathwohl, 2010; Ardahanlı, 2018; Bloom ve diğerleri, 1956; Bloom, Krathwohl ve Masia, 1984; Boekaert, Zeidner ve Pintich, 1999; Farr, 2010; Koray, Altunçekiç ve Yaman, 2005; Lewis ve Smith, 1993; Moon, 2004; Sosniak, 1994; Sosniak, 1994; Şahinel, 2002). Bu basamaklarda yer alan becerilerin göstergeleri öğretmen öz-yeterliğini gösterebilecek cümlelerle ifade edilmiştir. Ölçeğin varyansı açıklama oranı da göz önüne alındığında kapsam geçerliliğinin oldukça iyi olduğu sonucuna varılmıştır. Gerek ulusal gerekse uluslararası sınavlarda üst düzey öğrenme düzeyini ölçen sorular sorulmaktadır (Ar, 2019). Öğretmenlerin üst düzey öğrenme seviyesini ölçen soru hazırlama becerilerinin olması gerekmektedir. Bu becerilerin kazandırılmasına yönelik hizmet öncesi eğitim yer almazken hizmet sırasında alınabilecek ancak; pratikte yaygın hedefe ulaşması sadece bakanlık yoluyla zor olan eğitimler mevcuttur. Öğretmenlerin üst düzey öğrenme seviyesini ölçen soru hazırlama özelliklerinin olup olmadığının geçerli ve güvenilir bir araçla belirlenmesi gerekir. Bu anlamda ölçeğin alanyazına katkı sağlayabileceği düşünülebilir.

Sonuç olarak "Fen Bilimleri Öğretmenlerinin Üst Düzey Öğrenme Düzeyini Ölçen Soru Geliştirme Öz-yeterlik Ölçeği"nin fen bilimleri öğretmenlerinin üst düzey öğrenme düzeyini ölçen soru yazma öz-yeterliği seviyelerinin ortaya çıkarılmasında yararlanılabilecek geçerli ve güvenilir bir ölçek

olduğu sonucuna varılmış ve bu doğrultuda, ilgili ölçekten fen bilimleri öğretmenlerinin, fen eğitimi ve ölçme-değerlendirme alanında araştırmalar yürüten bilim insanlarının yararlanmaları önerilmektedir.


<http://kefad.ahievran.edu.tr>

Ahi Evran University

Journal of Kırşehir Education Faculty

ISSN: 2147 - 1037

ENGLISH VERSION

Introduction

Today's educational approaches are instead of individuals who take information without researching or questioning and only consuming them; It aims to raise individuals who research, question, know how to access information, produce information by applying the principles of access to information, and can apply it both to their own lives and to the use of humanity (Akpınar and Aydın, 2010). Accordingly, in new educational approaches, students should learn and internalize information instead of memorizing information. This case requires students to learn at high level (Doğanay, 1997). According to Bloom's Taxonomy, learning in the cognitive field takes place in six levels (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths and Wittrock, 2001). Knowledge, comprehension, application lower level, analysis, synthesis and evaluation are expressed as upper level (Şahinel, 2002). High-level learning and thinking complete each other. While low-level learning, previous learning and memorable information are expressed as repeating formulas, high-level learning guides students to analyze, interpret, and develop their ability to mobilize information (Anderson and Krathwohl, 2010). Instead of providing ready-made information, students should be directed to their ways of reaching information in order to lead them to higher-level learning. In this sense, it is beneficial to enable students to learn with real life tasks.

Lower and upper level learning is most commonly determined according to the revised Bloom Taxonomy. In this taxonomy, low-level learning includes remembering, understanding and application levels; high- level learning consists of analysis, evaluation and creation levels (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths and Wittrock, 2001; Lewis and Smith, 1993; Moon, 2004; Şahinel, 2002). Thinking skills are the indicator of learning skills. Whichever level of thinking the individual is at, he is also at that learning level (Güneş, 2012). The parsing level is the process of dividing the parts that make up a holistic information or a schema into parts as it is found in that whole and in the scheme. However, researching the consistency and validity links between the ideas put forward is within the scope of this level. The evaluation level can be described as the process of making a judgment by taking all circumstances into account. The creation level can be described as creating a whole, taking into account certain relationships and rules. But not all the products the student creates correspond to this concept (Bloom, Engelhart, Furst, Hill and Krathwohl, 1956; Bloom, Krathwohl and Masia, 1984;

Krathwohl, Bloom and Masia, 1964). There are situations such as innovation, originality, invention and creativity in creation (Simpson, 1966; Sosniak, 1994; Sönmez, 1993). It is important that the question items used in measurement tools are at the stage of analysis, evaluation and creation in order to determine whether the learning at this level is realized or not. Teachers who prepare measurement tools should have self-efficacy in preparing question items to measure high-level learning. Turkey is not readily available in a course or training program for senior learning directly measure spelling questions. Teacher candidates only take measurement and evaluation lessons in education during their undergraduate education. The content of this curriculum does not include an outcome for question writing that directly measures higher-level learning (YÖK, 2021). In-service training programs of teachers include a course for measuring high-level cognitive skills and preparing questions. In order to take this course, it is necessary to take two courses such as preparing open-ended questions 1-2 (MEB, 2021). It is very difficult for teachers to access this course due to the fact that in-service courses reach the widespread audience and they have to take three separate courses. It should be determined whether the teachers have self-efficacy regarding the questions they will prepare to measure higher-level learning. This can only be achieved with scales, which are psychometric measurement tools whose validity and reliability have been studied. When the literature is examined, it is seen that there is no such scale.

It is necessary to determine how much the high-level learning described above has been achieved. One of the most effective ways to determine how much higher-level learning has been achieved is to develop questions that measure higher-level learning. Bloom's Taxonomy is the most common form in which the cognitive level of the problem is determined in the question items (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths and Wittrock, 2001; Anderson and Krathwohl, 2010; Farr, 2010). A question item in the analysis level includes questions that want to measure the outlines of the issue, such as distinguishing the necessary and unnecessary data or establishing a hierarchy among the data on a topic (Bloom, Engelhart, Furst, Hill and Krathwohl, 1956; Lewis and Smith, 1993; Sosniak, 1994). The question items in the evaluation level are question items aiming to measure the gains that require checking the mistakes made in the solution of a problem and determining which of the multiple solutions will be more correct in solving the problem (Bloom, Engelhart, Furst, Hill and Krathwohl, 1956; Moon, 2004; Sosniak, 1994). The question items at the constitution level are question items that measure the ability to create alternative methods to these methods, although a previously unknown solution method needs to be developed or methods for its solution (Ardahanlı, 2018; Bloom, Engelhart, Furst, Hill and Krathwohl, 1956; Farr, 2010; Sosniak, 1994). In the analysis-level questions, the student's skills such as classification, description, grouping, comparison, differentiation, association, explanation are examined. In the evaluation-level questions, the student's skills such as reasoning, proposition, combining, planning, formulating, drawing conclusions, synthesizing are examined. In the questions at the constitution level, the student's skills such as decision making, judgment, preference, and proof are examined (Anderson, Krathwohl,

Airasian, Cruikshank, Mayer, Pintrich, Rath and Wittrock, 2001; Anderson and Krathwohl, 2010; Koray, Altunçekiç, and Yaman, 2005).

Education programs in the world in parallel with developments in Turkey are constantly renewed. In the last science program published in 2018, learning of metacognitive skills and the measurement and evaluation activities for them are emphasized (Ministry of National Education [MEB], 2018). Measurement and evaluation in education, by determining the level of acquisition of students; It gives feedback to educators in controlling and improving educational processes. Measurement and evaluation processes are carried out by means of the questions in measurement tools (Turgut & Baykul, 2015). Measurement and evaluation approach also changes depending on the changes in educational approaches. This causes the qualifications of the questions to change. While the questions used in exams until recent years measure students' recall and recognition of their knowledge; The questions in the exams prepared in recent years are aimed at measuring the level of analyzing, evaluating and forming students' existing knowledge (Doğan, 2019). Teachers prepare questions for both central exams and assessment and evaluation practices at schools. At this point, revealing the question-writing competencies of teachers becomes important in terms of the healthy conduct of measurement and evaluation activities in education.

When examining the literature of remembrance of science teachers in Turkey, while adequate comprehension and writing questions that measure learning at the application level; It is understood that it is insufficient in writing questions that measure the level of analysis, evaluation and creation that require high-level learning (Koray, Altunçekiç, and Yaman, 2005; Özcan and Akcan, 2010). A similar situation is valid for science teachers who are on active duty. The questions prepared by the science teachers on duty are mostly those that measure the lower level learning level (Akpınar and Ergin, 2006; Mutlu, Uşak and Aydoğdu, 2003; Cansüngü Koray and Yaman, 2002; Dindar and Demir, 2006; Ayvaci and Şahin, 2009) .

Studies determining the level of the questions prepared by science teacher candidates and teachers were determined as a result of the document analysis. In the determination process, the scale, which is an adapted psychometric tool or whose validity and reliability studies were conducted, was not used. Document reviews are weaker in terms of validity and reliability than screening patterns due to their nature (Cansız Aktaş, 2019: 127). At this point, there is a need for scales with validity and reliability studies to determine the self-efficacy levels of self-efficacy in question preparation that measure high-level learning.

Self-efficacy emerged with the socio-cognitive theory developed by Bandura (1977). Bandura (1977) defined the concept of self-efficacy as the belief of a person to accomplish a job at a certain level. On the other hand, self-efficacy is the belief in one's own abilities in order to take or organize the action required to demonstrate skills in any job (Bandura, 1997). In addition, teacher self-efficacy is the teacher's belief in the capacity to apply effective methods and techniques for teaching goals and

objectives (Park and Oliver 2008). According to the self-efficacy theory, high self-efficacy is an indicator of more determined teachers striving for the development of their students (Woolfolk, Winne, Perry and Shapka, 2009). Self-efficacy has two main components. These are the products of an action based on previous life experiences and their belief in their own abilities to solve the problems they encounter (Bandura, 1997).

Direct science teachers for in Turkey, STEM opinion scale (Gül Biçer, Uzoğlu of and Bozdoğan, 2018), related to the transition to secondary education system "science teacher opinion scale" (Taşkın and Aksoy, 2018), science teachers, laboratory and teaching process anxiety scale (Kahraman and Polat, 2017), opinion scale according to curriculum (Temli Durmuş and Ok, 2014), measurement and evaluation competency scale for primary school science and mathematics teachers (Yıldırım Ekinci and Köksal, 2011), personal professional development scale (Bilgin and Balbağ, 2016). On the other hand, it was understood that the scale of science teachers' attitudes about research and inquiry-based teaching (Açıkgöz, Uluçınar Sağır and Ozan, 2018) developed by Dibiase and McDonald (2015) was adapted.

The Aim of the Study

As a result of the literature review, it was understood that a limited number of scale development studies were conducted directly for science teachers (Biçer, Uzoğlu and Bozdoğan,2018; Durmuş Bakioğlu,2013; Taşkın and Aksoy,2018; Çevik,2017) . On the other hand, a scale developed solely for the measurement and evaluation of science teachers or writing questions was not encountered in the literature. It is thought that with this study, a scale will be added to the literature on assessment and evaluation self-efficacy for science teachers. Since scales are among psychometric measurement tools, their validity and reliability are affected by the characteristics of the study group (Büyüköztürk, 2012). Because of the thought that there would be a great difference in the self-efficacy perceptions of pre-service teachers and on-duty teachers, it was deemed appropriate by the researchers that the scale was intended only for science teachers. In this way, it is thought that scale development studies for science teachers will be enriched in terms of numerical and diversity. For this purpose, a scale development study was conducted to determine the science teachers' self-efficacy in developing high-level questions.

Method

This study is a quantitative scale development study designed as a descriptive survey model. Descriptive scanning model is a research method that aims to explain a past or present situation in the same way (Karasar, 2000). Since it is wanted to develop a scale that determines the self-efficacy of question development, which measures science teachers' high-level learning in this study, the model of the study is the scanning model.

Sample / Study Group / Participants

The universe of science teachers working in secondary schools are connected to the Ministry of Education in Turkey. The sample of the study consists of 518 science teachers who voluntarily participated. A total of 518 science teachers, 265 in the first application, 227 in the second application and 26 in the Test-Retest application contributed to the study. The distribution of some demographic characteristics of the teachers in the sample is presented in Table 1.

Table 1. *Descriptive distribution of the demographic characteristics of the teachers who make up the study group*

Specifications		1.		2.		Test Re -		Grand total	
		Application		Application		implementation			
		f	P(%)	f	P(%)	f	P(%)	f	P(%)
Gender	Male	121	45.66	103	45.37	16	61.50	240	46.33
	Female	144	54.34	124	54.63	10	38.50	278	53.67
	Total	265	100,00	227	100,00	26	100,00	518	100,00
Seniority	0-5 Years	52	19.62	38	16.74	0	0.00	90	17.37
	6-10 Years	82	30.94	76	33.48	6	23,10	164	31.66
	11-15 Years	59	22,26	53	23.35	9	34,60	121	23.36
	16-20 Years	31	11.70	22	9.69	4	15,40	57	11,00
	21-25 Years	25	9.43	17	7.49	5	19,20	47	9,07
	26 years and above	16	6.05	21	9,25	2	7.70	39	7.53
	Total	265	100,00	227	100,00	26	100,00	518	100,00
	Undergraduate	1	0.38	0	0.00	1	3.80	2nd	0.39
Education Level	License	198	74.72	181	79.74	18	69,20	397	76.64
	Post Graduate	61	23,01	39	17.18	6	23,10	106	20.46
	Doctorate	5	1.89	7	3.08	1	3.80	13	2,51
	Total	265	100,00	227	100,00	26	100,00	518	100,00
	Corum	26	9,81	25	11,01	18	69,20	69	13.32
City of Duty	Istanbul	22	8.30	27	11.89	0	0.00	49	9,46
	Sivas	36	13,59	5	2.21	0	0.00	41	7.92
	Amasya	21	7.92	13	5.73	3	11.50	37	7.14
	Other	160	60,38	157	69,16	5	19.30	322	62.16
	Total (68 Different Cities)	265	100,00	227	100,00	26	100,00	518	100,00

f: Frequency; P:Percent

When Table 1 examined; It is understood that male teachers are more than female teachers in terms of gender, but there is not much difference in proportion. When the working group is evaluated in terms of the seniority of the teachers; it is understood that the highest rate is teachers with a seniority of 6-10 years (31.66%) and the lowest rate is teachers with a seniority of 26 years or more (7.53%). In terms of education level, it is understood that the majority of them are undergraduate (76.64%). When evaluated in terms of the cities where the working group worked, it is understood that the highest attendance was from Çorum, Istanbul, Sivas and Amasya, respectively. In addition, there are science teachers from 68 different cities in the study group.

Data Collection Tools / Data Collection Methods / Creating Item Pool

During the development of the scale, the literature was first reviewed. In this framework, studies investigating the assessment and evaluation of science teachers were examined. As a result of this examination, it was determined that science teachers need to demonstrate the self-efficacy of developing high-level questions. Later, studies on the concept of self-efficacy were examined. Thus, the background of the concept of self-efficacy was determined (Bandura, 1977; Park and Oliver 2008; Woolfolk, Winne, Perry and Shapka, 2009). How to create high-level learning and thinking concepts and indicators was determined by a literature review (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths and Wittrock, 2001). Based on the information obtained, high-level learning consists of analysis, evaluation and creation levels of the cognitive process steps included in the Bloom taxonomy, and therefore, the study of creating an item pool for self-efficacy measuring learning outcomes in three levels by researchers. In the first stage, an item pool consisting of 46 items was created. In the item pool, there were 19 items for the analysis level, 15 items for the evaluation level and 12 items for the creation level. While creating the items, the characteristics of the questions in the analysis, evaluation and creation levels in the Bloom Taxonomy were investigated. As a result of the research, it was determined which skills and acquisitions the questions in these three levels examine in students. Classification, combining, interpretation, association, suggestion, determining variables, measuring etc. belonging to these skills and acquisitions. Keywords were determined as indicators of high level learning for scale items. In this framework, an item pool was created. The created item pool was arranged in five-point Likert type to determine the self-efficacy levels of science teachers specified in the items. These options are; it has been prepared and rated as strongly disagree (1), disagree (2), undecided (3), agree (4) and strongly agree (5). Two science education specialists, two linguists, one educational science expert and two assessment and evaluation experts were presented to the view and the validity of the content was tried to be achieved. As a result of expert opinion, 3 items were excluded from the scale on the grounds that they were items measuring similar characteristics. As a result of the recommendation of a science education expert, 10 items were added to the scale. The necessary changes were made in line with the opinions of the language and assessment and evaluation experts and the 53-item scale was made ready for pilot application. The pilot application of the scale was carried out with 5 science teachers. It was paid attention that the teachers in the pilot application were the group that reflected the sample, and the results of the pilot applications were not included in the main analysis. In line with the warnings of the teachers, although the content remained the same, it was made ready for the main application by making arrangements in terms of language and semantic knowledge.

Analysis of Data

Based on the data obtained for the scale within the scope of statistical analysis, KMO and Bartlett test analyzes were performed first to determine whether it is suitable of the data to factor analysis in order to examine the construct validity of the scale. As a result of the analysis, because the data were

suitable for factor analysis, exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) were performed on the data; The distribution of scale items to factors was determined by principal component analysis; Factor loadings were determined with Varimax Vertical Rotation technique. EFA is a type of statistical analysis that provides information about the structure of the measurement tool to be developed and explains the variables that measure the same structure with as few factors as possible (Balci, 2013). As a result of Principal Component Analysis, items with factor loadings less than 0.30 and items with a load of two or more factors and a difference of 0.10 or less between the load values taken should be excluded from the scale (Büyüköztürk, 2012). However, if the factor load value is below 0.40 and exclusion from the scale does not significantly affect the reliability of the scale; Removing these items can make the scale more valid and reliable. It is sufficient for behavioral sciences that the factor load values of the scale items are greater than 0.30 and the general variance explains at least 30% in single- factor scales and at least 40% in multi-factor scales (Kline, 1994; Scherer, Wiebe, Luther. and Adams, 1988). When evaluating factor analysis results, the main criterion is factor loadings (Gorsuch, 1983; Balci, 2013). The high factor load values indicate that the variable can be found under the relevant factor (Büyüköztürk, 2012). In addition, it is important to calculate the common factor variance for multi-factor scales, and with principal component analysis, it is expressed as the common variance caused by factors on each variable (Çokluk, Şekercioğlu and Büyüköztürk, 2010). The common factor variance is expected to be greater than 0.50 (Erkuş, 2016). In this context, items with a common factor variance value less than 0.50, factor load less than 0.40, and a difference of 0.10 or less between factor loads by distributing to more than one factor were removed from the scale by consulting experts' opinions.

The scale form that emerged as a result of the exploratory factor analysis was applied to a different study group from the study group of the first application, and confirmatory factor analysis was performed on the data obtained. Confirmatory Factor Analysis is the analysis method based on the principle of evaluating and testing the relationships between observed and latent variables as hypotheses (Pohlmann, 2004). The maximum likelihood technique was used in the confirmatory factor analysis. The maximum likelihood technique was preferred because it is both flexible and easy to understand. The observed values in the scale model formed by confirmatory factor analysis are expected to be within the ranges specified in Table 2 (Kline, 2005; Sümer, 2000; Şimsek, 2007).

Table 2. *The ranges of fit indices of the values observed in the scale model formed by confirmatory factor analysis*

Harmony Indices	Perfect Match	Acceptable Compliance
χ^2 / d	$\chi^2 / d < 3$	$\chi^2 / d < 5$
RMSEA	$0 < RMSEA < 0.05$	$0.06 \leq RMSEA < 0.08$
S- RMR	$0 \leq S- RMR \leq 0.05$	$0.06 \leq S- RMR \leq 0.08$
NNFI	$0.97 \leq NNFI \leq 1$	$0.90 \leq NNFI \leq 0.96$
CFI	$0.97 \leq CFI \leq 1$	$0.90 \leq CFI \leq 0.96$
GFI	$0.95 \leq GFI \leq 1$	$0.90 \leq GFI \leq 0.96$
AGFI	$0.95 \leq AGFI \leq 1$	$0.90 \leq AGFI \leq 0.96$
IFI	$0.95 \leq IFI \leq 1.00$	$0.90 \leq IFI \leq 0.95$

The item discrimination of the 30-item scale obtained by factor analysis was analyzed using the t test for independent samples. The level of each item of the scale serving the general purpose of the scale and item-total correlations were analyzed with Pearson's r test. In single-factor scales, the correlation between the score calculated separately from all items and the score calculated from the scale in general is at a positive and significant level, and it is used as a criterion in terms of determining the level of the items of the scale separately, and therefore the level of serving the general purpose of the scale (Balci, 2013). Internal consistency coefficients and stability coefficients were calculated to determine the reliability of the scale. Cronbach alpha reliability coefficient, Guttman split-half, two-halves correlation coefficient and Sperman-Brown coefficients were used to calculate the internal consistency level. A reliability coefficient of at least 0.70 is an indication that the scale is reliable (Gorsuch, 1983; Büyüköztürk, 2002). The stability level of the scale was calculated by calculating the correlation between the results of the two applications performed 30 days apart, and the relationship between the two applications. In order for a scale to be stable, there should be acceptable consistency between the results of two applications performed at regular intervals. The reliability coefficient, which expresses the consistency level, varies between 0 and 1, and the closer it gets to 1, the higher the consistency level (Gorsuch, 1983). For correlation coefficients, $r \leq 0,30$ indicates low level, $r \leq 0,70$ medium level, and $r \leq 1,00$ level indicates high correlation (Büyüköztürk, 2002).

Ethical Issues

In this study, all rules stated to be followed within the scope of "Higher Education Institutions Scientific Research and Publication Ethics Directive" were followed. None of the actions stated under the title "Actions Against Scientific Research and Publication Ethics", which is the second part of the directive, were not taken. Ethics committee approval, the information of which was shown below, was obtained prior to the study.

Ethics committee permission information: Name of the board conducting the ethical review: Amasya University Social Sciences Ethics Committee

Date of the ethical assessment decision :15.05.2020

Ethics assessment document issue number :30640013-044

Findings

In this part of the study, the operations and findings within the scope of validity and reliability analysis of the scale are included.

Findings Regarding the Validity of the Scale

In this part of the study, the construct validity, item discrimination and item-total correlations were analyzed in order to determine the validity of the question development self-efficacy scale measuring the high-level learning level of science teachers, and the findings were presented:

Structure validity

Findings Regarding Exploratory Factor Analysis: In order to determine the construct validity of the scale, the data were first subjected to Kaiser-Meyer-Okin (KMO) and Bartlett test analyzes and KMO = 0,966; Bartlett test value is $\chi^2 = 14150,785$; $sd = 1378$ ($p = 0.000$). Within the scope of these results, it was understood that the data of the 53-item scale were suitable for factor analysis.

Upon understanding that the scale is suitable for factor analysis, principal component analysis was conducted to determine how many dimensions the scale is. Principal component analysis is one of the most used techniques among factoring techniques (Büyüköztürk, 2002). Afterwards, Varimax Vertical Rotation Technique was used in terms of its basic components. Accordingly, a total of 11 items, 1 with a factor load of less than 0.40 and 10 with a factor load of less than 0.1, were removed from the scale, and a re-factor analysis was performed on the remaining 42 items. As a result of the second analysis, it was understood that the remaining 42 items in the scale were clustered under one factor. Thereupon, re-exploratory factor analysis was performed on 42 items for the third time as a one-factor scale structure. As a result of these analyzes, 8 items with a common variance value of less than 0.50 and 4 more items with a factor load of less than 0.40 were removed from the scale. So; A total of 23 items were dropped from the scale, and a 30-item factorial scale was obtained in the last case. In order to ensure that the discarded items do not spoil the content validity, the new item pool was examined by two science education experts again. These two experts are the experts who made the first evaluation. Following the opinions of science education experts that excluding 23 items did not affect the content validity negatively, other analyzes were carried out. In the last case, the KMO value of the 30-item scale was 0.967; Bartlett Test value $\chi^2 = 8057,031$; $sd = 435$; It was calculated as $p = 0.00$. The factor loads of the remaining 30 items in the scale were determined to be between 0.713 and 0.856 without being subject to rotation. On the other hand, it was determined that the items and single factor within the scope of the scale explained 63,861% of the total variance. When the literature is examined, it is considered sufficient that the factor loads cannot be less than 0.30 and the variance percentage explained in social sciences is at least 30% in single- factor structures and at least 40% in multi-factor structures (Büyüköztürk, 2012; Eroğlu, 2008). Due to the structure of the scale, the single factor with 30 items is named as "science teachers' high-level question development self-efficacy scale."

The situation explained above is presented in Graph 1 in the slope-accumulation graph drawn according to the eigenvalues of the factors.

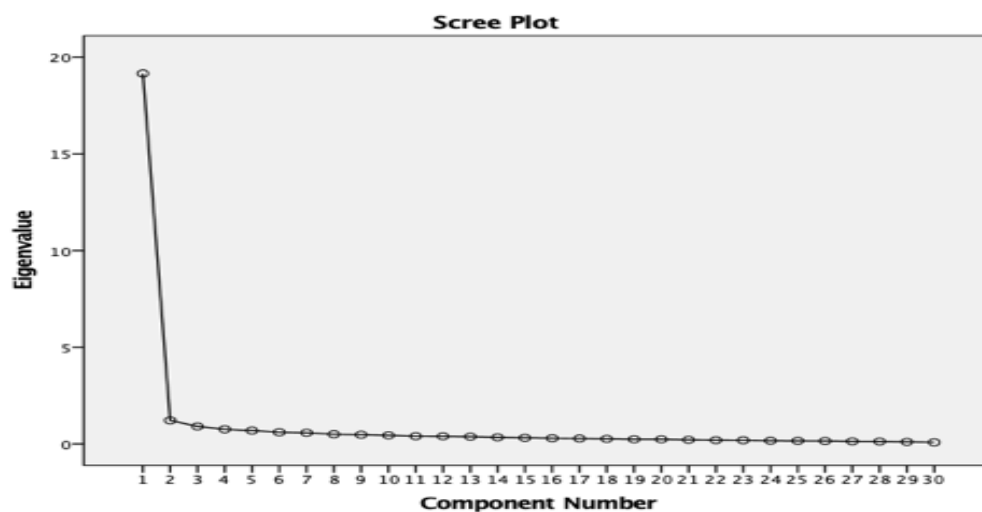


Chart 1. Slope plot chart (eigenvalues by factors)

Graph 1 shows that there is a high accelerated decline after the first factor; For this reason, a factor contributes to the variance at a high rate; On the other hand, it is understood that other factors contribute little to the variance compared to the first factor, the lines begin to take a horizontal image after the first breaking point, in other words, their contribution to the variance is close to each other (Büyüköztürk, 2012; Eroğlu, 2008).

Findings regarding the item loadings of the remaining 30 items in the scale in terms of single factor and the factor's eigenvalue and variance explanation ratio Table 3 It is specified in.

Table 3. Factor analysis results of the scale based on a single factor

Factor	Item No	Common Variances	Common Factor Variance	Factor Load	Item No	Common Variances	Common Factor Variance	Factor Load
Self-Efficacy Assessing High Level Learning	I1	0.585	0.585	0.856	I16	0,689	0,689	0,809
	I2	0,678	0,678	0.848	I17	0,650	0,650	0.806
	I3	0.666	0.666	0.845	I18	0,698	0,698	0.797
	I4	0.732	0.732	0.842	I19	0.532	0.532	0.786
	I5	0,662	0,662	0.840	I20	0,657	0,657	0.783
	I6	0.606	0.606	0.837	I21	0.714	0.714	0.778
	I7	0,613	0,613	0.835	I22	0.584	0.584	0.771
	I8	0,701	0,701	0.833	I23	0.594	0.594	0.765
	I9	0,655	0,655	0.830	I24	0.575	0.575	0.764
	I10	0.635	0.635	0.828	I25	0,509	0,509	0.758
	I11	0,709	0,709	0.825	I26	0,686	0,686	0.745
	I12	0.719	0.719	0.823	I27	0,617	0,617	0.739
	I13	0,705	0,705	0.816	I28	0.547	0.547	0.730
	I14	0,681	0,681	0.814	I29	0.511	0.511	0.715
	I15	0,694	0,694	0.810	I30	0.556	0.556	0.713
				Eigenvalue		19.158		
				Variance Explained		63,861		

Table 3 when examined, the scale included 30 items and a single factor; It is seen that the factor loads of the items vary between 0.713 and 0.856. The eigenvalue of this factor on the scale is 19,158; The

rate of contribution to variance is 63,861%. In addition, it is seen that the common factor variance value of all items is higher than 0.50.

Confirmatory Factor Analysis on the Results: After the exploratory factor analysis, confirmatory factor analysis was performed on the data collected from 227 science teachers, a different sample, in order to verify the factor structures of the scale, which was determined to consist of a single factor. The model obtained as a result of confirmatory factor analysis performed using the maximum likelihood technique without any limitations is presented in Figure 1.

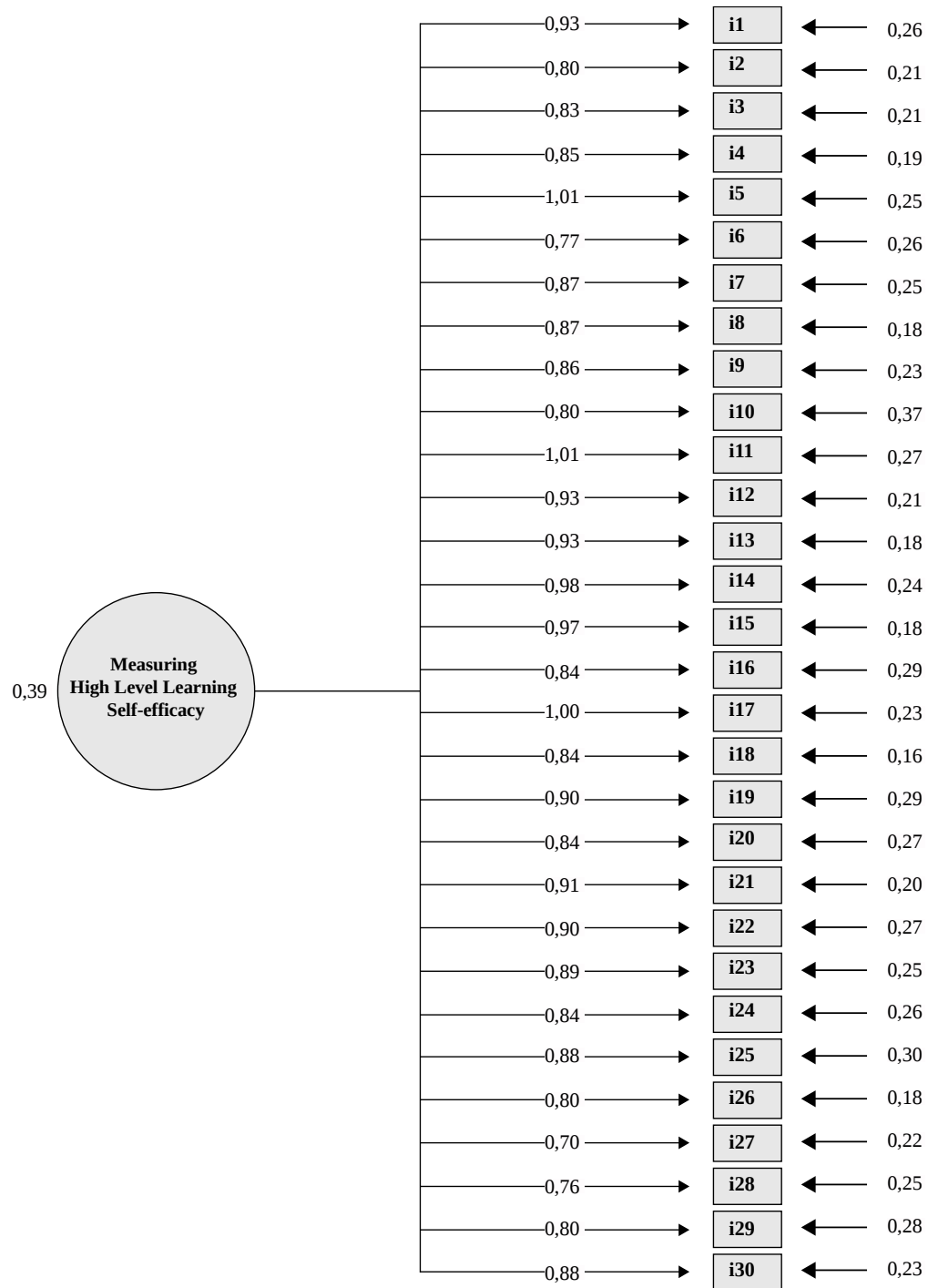


Figure 1. Confirmatory factor analysis connection diagram of the scale

In the path diagram plotted in Figure 1, standardized values indicating how much each item represents the factor and hence the scale. When the model examined, it was found that the error variance of the items varied between 0.18 and 0.30; It is understood that the structural coefficient between the latent variable and the observed variable varies between 0.70 and 1.01.

The values of the goodness of fit index calculated as a result of the confirmatory factor analysis Table 4. presented in.

Table 4. *Fit index values obtained as a result of confirmatory factor analysis*

Harmony Indices	Index Values
χ^2 / d	2,657
RMSEA	0.840
S- RMR	0.025
NNFI	0.828
CFI	0.884
GFI	0.768
AGFI	0,729
IFI	0.885

When Table 4 examined, it is calculated as $\chi^2 / d = 2.657$, RMSEA = 0.840, S-RMR = 0.025, NNFI = 0.828, CFI = 0.884, GFI = 0.768, AGFI = 0.729 and IFI = 0.885. When the goodness of fit indices are evaluated, it can be said that the model is χ^2 / d good fit, RMSEA acceptable fit and S-RMR value within perfect fit intervals. NNFI, CFI, GFI, AGFI and IFI goodness of fit indices are not within acceptable range of values. However, at least three goodness of fit index values are sufficient to validate the model by the data (McDonald and Ho, 2002). This results shows that the model is verified by the data.

Item Discrimination: The distinctiveness of the items in the scale and the entire scale was examined. In this context, it was examined whether there was a significant difference between the 27% lower and upper groups of the scores calculated from the entire scale. For this purpose, independent samples t-test was conducted based on the scores of each item and total scores. Analysis results are presented in Table 5.

Table 5. Independent sample t test results for the item discrimination features of the scale

Item Number	Average		t	Item Number	Average		t
	Subgroup	Upper Group			Subgroup	Upper Group	
i1	3.25	4.54	-14,924	i16	3.39	4.63	-14,383
i2	3,52	4.74	-14,436	i17	3,1 0	4.56	-17,557
i3	3.48	4.71	-14,538	i18	3.47	4.66	-15,366
i4	3.44	4.68	-15,033	i19	3.18	4.44	-13,478
i5	3.13	4.57	-16,947	i20	3.35	4.57	-13,691
i6	3.65	4.8 0	-12,549	i21	3.28	4.56	-15,212
i7	3.32	4.56	-14,316	i22	3.06	4.39	-14,534
i8	3.32	4,59	-14,716	i23	3.29	4.53	-14,606
i9	3,58	4.8 0	-14,158	i24	3.21	4.44	-14,171
i10	3.65	4.79	-11,953	i25	3.12	4.38	-14,069
i11	3.2 0	4.66	-16,616	i26	3.38	4.5 0	-14,238
i12	3.21	4,59	-16,481	i27	3.35	4.45	-13,663
i13	3.33	4,6 0	-15,161	i28	3.45	4.47	-12,076
i14	3.23	4,458	-16,393	i29	3.49	4.61	-13,729
i15	3.33	4.61	-15,108	i30	3.11	4.47	-17,223
Total				105.85	146,338	-22.76	

N = 492; ** = $p < 0.05$

When table 5 examined, it is understood that there is a significant difference ($p < 0.05$) between the lower and upper group scores of all items and the scale in general. This situation shows that the general and all of the items of the scale have separate distinguishing features.

In addition, according to the item-total correlation technique, the level of each item serving the general purpose was determined by calculating the correlations between the scores obtained from the overall scale, and each item separately. Item-factor correlation values calculated for each item Table 6. presented in.

Table 6. Item-factor (total) scores correlation analysis

Item Number	r	Item Number	r
i1	0.761 *	i16	0.778 *
i2	0.788 *	i17	0.802 *
i3	0.789 *	i18	0.818 *
i4	0.818 *	i19	0.734 *
i5	0.799 *	i20	0.775 *
i6	0.743 *	i21	0.818 *
i7	0.769 *	i22	0.757 *
i8	0.816 *	i23	0.765 *
i9	0.782 *	i24	0.749 *
i10	0.734 *	i25	0.723 *
i11	0.820 *	i26	0.801 *
i12	0.823 *	i27	0.740 *
i13	0.821 *	i28	0.720 *
i14	0.809 *	i29	0.701 *
i15	0.826 *	i30	0.753 *

N = 492; * = $p < 0.001$

When Table 6 examined, it is understood that the correlation values of the items vary between 0.701 and 0.826, and all relationships are positive and significant. This situation shows that each item is valid and consistent with the overall scale, thus serving the general purpose of the scale.

Findings Regarding the Reliability of the Scale

Internal consistency and stability analyzes were performed on the data to determine the reliability of the scale. The findings obtained with these processes are shown below:

Internal consistency level: Reliability analysis of the scale; The correlation value between two identical halves, Cronbach alpha reliability coefficient was calculated using Guttman split-half and Sperman-Brown formulas. The findings obtained from these tests for the reliability of the scale are presented in Table 7.

Table 7. *Reliability analysis results of the scale*

Factor (Scale)	N	Number of Items	Two-Peer-Half Correlations	Sperman Brown	Guttman Split-Half	Cronbach Alpha
Self-Efficacy Assessing High Level Learning	492	32	0,902	0.948	0.948	0.977

Table 7 when analyzed, the two-half correlations of the scale consisting of one factor and 30 items are 0,902; Sperman Brown reliability coefficient is 0.948; Guttman Split-Half value of 0.948; The Cronbach alpha reliability coefficient was calculated as 0.977.

Stability level: One of the features of a reliable measurement tool is stability. It is emphasized that there should be a positive and significant relationship between the two applications when a stable measurement tool is applied to the same sample group for a certain time, for example, at least three weeks apart (Balci, 2013). The stability level of the scale was determined through the test-retest method. The 30-item final version of the scale was administered to 26 teachers twice, 30 days apart. The consistency between the scores calculated as a result of the two applications was examined both separately for all items and in terms of the overall scale. In this way, the stability of each item and the entire scale in the scale was tested. Findings obtained were presented in Table 8.

Table 8. *Test-retest results of the items of the scale*

Item Number	r	Item Number	r
i1	0.641 **	i16	0.772 **
i2	0.687 **	i17	0.545 **
i3	0.627 **	i18	0.770 **
i4	0.725 **	i19	0.587 **
i5	0.393 *	i20	0.725 **
i6	0.735 **	i21	0.657 **
i7	0.656 **	i22	0.547 **
i8	0.449 *	i23	0.556 **
i9	0.548 **	i24	0.715 **
i10	0.836 **	i25	0.700 **
i11	0.478 **	i26	0.489 **
i12	0.637 *	i27	0.589 **
i13	0.806 **	i28	0.700 **
i14	0.380	i29	0.773 **
i15	0,643 **	i30	0.613 **
Total		0.833 *	

N: 26; * = $p < 0.05$; ** = $p < 0.01$

When Table 8. examined, it was observed that all of the items that make up the scale, the correlation coefficients calculated by the test-retest method varied between 0.380 and 0.836, and the relationship for each item was significant and positive except for the 14th item ($p < 0.01$ or $p < 0.05$). Although the correlation of item 14 was positive, no significant correlation could be determined. However, in line with the expert opinions, it was decided to keep the item in the scale due to the feedback that this item is important for content validity. The correlation coefficient for the whole scale was positive and significant in the finding obtained for the stability level of the whole scale ($r = 0.833$ and $p < 0.05$). A reliable measuring tool; must be stable, consistent and responsive. In this sense, the values calculated in the form of stability coefficient can be evaluated as an indicator of the reliability of the scale (Hovardaoğlu, 2000). When the scale developed in this respect is evaluated, it can be said that the scale has the desired feature.

Conclusions, Discussion and Suggestions

In this study, carried out in order to develop a question development self-efficacy scale measuring high level learning of science teachers, the relevant scale is valid and highly reliable, Likert type, single factor and consists of 30 items. All of the items; It is rated as Strongly Disagree (1), Disagree (2), Undecided (3), Agree (4), Strongly Agree (5). The KMO value of the 30-item scale is 0.967; Bartlett Test values $\chi^2 = 8057,031$; $sd = 435$; $p = 0.00$. The factor loads of 30 items of the scale are between 0.713 and 0.856 without being subject to rotation (unrotated). The items and single factor in the scale explain 63,861% of the total variance. In addition, the common factor variance value of all items is greater than 0.50. The fit index values calculated as a result of the confirmatory factor analysis of the scale are $\chi^2 / d = 2.657$, RMSEA = 0.840, S-RMR = 0.025, NNFI = 0.828, CFI = 0.884, GFI = 0.768, AGFI = 0.729 and IFI = 0.885. When the goodness of fit indices are evaluated, it is understood that the model of the scale is verified by the data. Because there are many types of fit indices for the model to be verified by the data.

It is sufficient for at least three of these indices to be at an acceptable level (McDonald and Ho, 2002). As a result of the item analysis of the scale, there is a significant difference ($p < 0.05$) between the lower and upper group scores of all items and the overall scale. For this reason, the overall scale and all items have distinctive features (Balci, 2013; Erkuş, 2016). According to the item-total correlation analysis of the scale, the correlation values of the items vary between 0.701 and 0.826, and all relationships are positive and significant. Therefore, each item is valid and consistent with the overall scale, so each item serves the general purpose of the scale. The two halves correlations of the scale are 0,902; Spearman Brown reliability coefficient is 0.948; Guttman Split-Half value of 0.948; The Cronbach Alpha reliability coefficient is 0.977. For this reason, the reliability of the scale is very high (Büyüköztürk, 2002). The correlation coefficients of all items of the scale calculated by test-retest method are between 0.380 and 0.836, and the relationship for each item is significant and positive except for item 14 ($p < 0.01$ or $p < 0.05$). However, in line with expert opinions, article 14 is included in the scale as it is important for content validity. The correlation coefficient for the stability level of the whole scale is positive and significant ($r = 0.833$ and $p < 0.05$). When the psychometric properties of the scale are evaluated together, it can be concluded that the scale is valid and reliable.

When the literature examined, in the last decade, it is seen that there are scales such as the scale of measurement and evaluation competencies of teachers who teach science and mathematics at grade level 4-8 at secondary school (Yıldırım Ekinci and Köksal, 2011) and the scale of measurement and evaluation competencies of teachers (Kuzu, 2016). However, no scale dealing with the question writing skills of teachers was found. It was concluded that the scale developed in this study is important in determining the self-efficacy of self-efficacy for self-efficacy of science teachers in measuring high-level learning. The validity feature of the scale was examined by three different analyzes. These are factor analysis, item-total correlation analysis, and t-test analyzes for independent samples analyzing the discrimination characteristics of the scale between 27% subgroup and supergroup. When scale development studies are examined, the most commonly used technique among factor analytical methods is principal component analysis. In scale development studies in the literature, construct validity is only made through exploratory factor analysis (Erkuş, 2016). In this study, with the analysis of principal components within the scope of exploratory factor analysis in accordance with the literature; It is understood that the scale consists of a single factor. When the factor loads of each item, the eigenvalue of the factor and the explained variance ratio are examined, it can be stated that the scale is a scale with construct validity.

When the literature examined, it is seen that most of the scale development studies examine the construct validity only through exploratory factor analysis (Gül and Sözbilir, 2015). According to Gül and Sözbilir (2015), the reason for this situation is; While it can be done with widely known and used software such as EFA's SPSS; The fact that CFA can be performed in less known and uncommon programs such as LISREL and Amos, therefore, is due to the lack of knowledge and experience. EFA is

aimed at understanding the structure of the psychometric measurement tool developed, and CFA is aimed at testing this defined structure. Both analyzes are important at the stage of scale development and are the type of analysis that do not complement each other. For this reason, the structure of the scale should be determined with EFA and then the structure defined by CFA should be tested in the scale development phase (Erkuş, 2016). In this study, confirmatory factor analysis was used to confirm the factor structure of the scale, which was determined to consist of a single factor by exploratory factor analysis. Among the fit indices calculated by confirmatory factor analysis, there are three values: good, excellent and acceptable. If at least three of the fit indices are at an acceptable or better level, it means that the data are verified (McDonald and Ho, 2002). In this case, it shows that the results of the confirmatory factor analysis and the observed value of the scale model are compatible with the data, in other words, this model is verified by the data.

Item-total correlations were calculated on the data to determine the level at which all items in the scale could measure the features that were tried to be measured with the scale. According to the calculated values, the correlation of each item with the total score is positive and significant; It has been concluded that each item in the scale serves the goal of measuring the feature desired to be measured with the entire scale. Item analysis is also very important at the stage of scale development, and the distinctiveness of the items should be examined (Erkuş, 2016). Therefore, in this study, whether the items are discriminatory or not was tested. With the t-test for independent samples, it was concluded that each item and the general of the scale had a distinctive feature. In scale development studies prepared in the literature, mostly item-test correlation and item analysis are performed (Şahin and Boztunç Öztürk, 2018; Gül and Sözbilir, 2015). The literature was scanned to determine the criterion validity of the scale; criterion validity could not be determined since there was no similar scale in terms of content and target.

The most frequently preferred analysis for reliability calculations in scale development studies is the Cronbach α test (Gül and Sözbilir, 2015; Tosun and Taşkesenligil, 2015). In this study, the internal consistency coefficients of the scale; Cronbach Alpha was calculated with Sperman-Brown formula, two-half correlations and Guttman split-half reliability formula. Considering the values of these calculations, it can be concluded that the reliability of the scale is very high and it can make reliable measurements.

In order to determine the stability of the scale, the test-retest method and correlation analysis between the two applications were performed on the data obtained with the applications performed 30 days apart. These analyzes were calculated for all items and for the entire scale. According to the calculated correlation coefficients, there is no item with a low level of correlation among the items in the scale. Of the items, 21 had moderate and 9 had a high level of correlation. Also, the general correlation of the scale is high. All items except item 14 and the overall scale have positive (including item 14) and significant correlations. In this case, it has been concluded that the scale can make stable

measurements. In prepared scale development study in Turkey are very few stability analysis (Şahin and Boztunç Öztürk, 2018).

When the items in the scale evaluated, the items measure the teachers' self-efficacy in developing questions for the analysis, evaluation and creation levels in Bloom Taxonomy. These levels in Bloom's Taxonomy were taken into account while creating the scale item pool. Because these levels represent high-level learning levels with common acceptance (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths and Wittrock, 2001; Anderson and Krathwohl, 2010; Farr, 2010). The scale items were prepared in line with expert opinions, taking into account the indicators of the skills in the analysis, evaluation and creation levels in the Bloom Taxonomy. The items in the analysis level examine the skills of classifying, supporting, collecting evidence, interpreting, making inferences, establishing relationships, dividing into parts, making inferences, collecting data, saving data, analyzing data, and interpreting (Koray, Altunçekiç ve Yaman, 2005). While preparing the items in the evaluation level, comparing, proving, predicting, criticizing, measuring, predicting, and decision-making skills are examined. On the other hand, the items in the creation level examine the skills of proving, proposing, organizing, organizing, combining, developing, designing, proving, presenting and sharing. Students with high-level learning should have these skills (Anderson, Krathwohl, Airasian, Cruikshank, Mayer, Pintrich, Raths and Wittrock, 2001; Anderson and Krathwohl, 2010; Ardahanlı, 2018; Bloom, Engelhart, Furst, Hill, and Krathwohl, 1956. ; Bloom, Krathwohl and Masia, 1984; Boekaert, Zeidner and Pintich, 1999; Farr, 2010; Koray, Altunçekiç, and Yaman, 2005; Lewis and Smith, 1993; Moon, 2004; Sosniak, 1994; Sosniak, 1994; Şahinel, 2002). The indicators of the skills in these levels are expressed in sentences that can show teacher self-efficacy. Considering the variance explanation ratio of the scale, it was concluded that the content validity was quite good. In both national and international exams, questions measuring high-level learning are asked (Ar, 2019). Teachers should have question preparation skills that measure the high level of learning. Pre-service training for the acquisition of these skills is not included, but can be taken during the service; there are trainings that are difficult to achieve in practice only through the ministry. It should be determined with a valid and reliable tool whether there are question preparation features that measure teachers' high level learning. In this sense, it can be thought that the scale can contribute to the literature.

As a result, it has been concluded that the Question Development Self-Efficacy Scale Measuring the High-Level Learning Level of Science Teachers is a valid and reliable scale that can be used to reveal the self-efficacy levels of self-efficacy in question writing, and accordingly, It is recommended that science teachers, scientists who conduct research in the field of science education and assessment and evaluation benefit.

References

- Açıkgöz, D., Sağır, Ş. U., & Ozan, F.(2018). Araştırma sorgulamaya dayalı öğretim hakkında öğretmen tutumları ölçek uyarlama çalışması. *Uluslararası Türk Eğitim Bilimleri Dergisi*, 6(11), 1-14.
- Akpınar, E., & Ergin, Ö. (2006). Fen bilgisi öğretmenlerinin yazılı sınav sorularının değerlendirilmesi. *Milli Eğitim Dergisi*, 35(172), 225-231.
- Akpınar, B., & Aydın, K. (2010). Eğitimde değişim ve öğretmenlerin değişim algıları. *Eğitim ve Bilim*, 32(144), 71-80.
- Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., Raths, J., & Wittrock, M.C. (2001). The Revised Taxonomy Structure. In. Anderson, L.W. (Ed.), Krathwohl, D.R. (Ed.), *A taxonomy for learning, teaching, and assessing: A revision of Bloom's Taxonomy of Educational Objectives* (Complete edition). New York: Longman.
- Anderson, L. W., & Krathwohl, D. R. (2010). Bloom'un eğitim hedefleri ile ilgili sınıflamasının güncelleştirilmiş biçimi. *Durmuş Ali Özçelik (Çev.). Ankara: Pegem Akademi*.
- Ar, M. E. (2019). *Fen bilimleri öğretmenlerine yönelik geliştirilen nitelikli yaşam temelli açık uçlu soru hazırlama kursunun uygulanması ve değerlendirilmesi*. Yayınlanmamış Yüksek Lisans Tezi, Uludağ Üniversitesi Eğitim Bilimleri Enstitüsü, Bursa.
- Ardahanlı, Ö. (2018). *TEOG sınavı matematik soruları ile 8. sınıf matematik yazılı sınav sorularının yenilenmiş bloom taksonomisine göre incelenmesi*. Yayınlanmamış Yüksek Lisans Tezi, Osmangazi Üniversitesi Eğitim Bilimleri Enstitüsü, Eskişehir.
- Ayvacı, H. Ş., & Türkdoğan, A. (2010). Yeniden yapılandırılan bloom taksonomisine göre fen ve teknoloji dersi yazılı sorularının incelenmesi. *Türk Fen Eğitimi Dergisi*, 7(1), 13-25.
- Balcı, A. (2013). *Sosyal bilimlerde araştırma: yöntem teknik ve ilkeler*. (10. Basım). Ankara: PegemA Yayıncılık.
- Bandura, A. (1997). *Self-efficacy: the exercise of control*. New York, NY: W.H. Freeman and Co.
- Biçer, B. G., Uzoğlu, M., & Bozdoğan, A. E. (2018). Fen bilimleri öğretmenlerinin stem hakkındaki görüşlerinin belirlenmesine yönelik ölçek geliştirme çalışması. *OPUS Uluslararası Toplum Araştırmaları Dergisi*, 9(16), 551-574.
- Bilgin, A., & Balbag, M. Z. (2016). Personal professional development efforts scale for science and technology teachers regarding their fields. *Acta Didactica Napocensia*, 9(2), 67-78.
- Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., & Krathwohl, D.R. (1956). The Taxonomy and Illustrative Materials. In. Bloom, B.S. (Ed.), *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain*. New York: David McKay.

- Bloom, B. S., Krathwohl, D. R., & Masia, B. B. (1984). *Bloom taxonomy of educational objectives*. In Allyn and Bacon. Pearson Education.
- Büyükoztürk, Ş. (2012). *Sosyal bilimler için veri analizi el kitabı*. (16.Basım). Ankara: PegemA Yayıncılık.
- Cansız Aktaş, M. (2019). Nitel veri toplama teknikleri. H.Özmen ve O. Karamustafaoğlu içinde, *Eğitimde Araştırma Yöntemleri* (113-136), Ankara: Pegem Akademi.
- Cansüngü Koray, Ö., & Yaman, S. (2002). Fen bilgisi öğretmenlerinin soru sorma becerilerinin bloom taksonomisine göre değerlendirilmesi. *Kastamonu Eğitim Dergisi*, 10(2), 317-324.
- Çevik, M. (2017). Ortaöğretim öğretmenlerine yönelik FeTeMM farkındalık ölçeği (FFÖ) geliştirme çalışması. *Journal of Human Sciences*, 14(3), 2436-2452. doi:10.14687/jhs.v14i3.4673
- Çokluk, Ö., Şekercioğlu, G., & Büyükoztürk, Ş. (2010). *Sosyal bilimler için çok değişkenli istatistik: Spss ve Lisrel uygulamaları*. Ankara: PegemA Yayıncılık.
- Dibiase, W. & McDonald, J.R. (2015). Science teacher attitudes toward inquiry-based teaching and learning. *The Clearing House*, 88: 29–38. DOI: 10.1080/00098655.2014.987717
- Dindar, H., & Demir, M. (2006). Beşinci sınıf öğretmenlerinin fen bilgisi dersi sınav sorularının bloom taksonomisine göre değerlendirilmesi. *Gazi Eğitim Fakültesi Dergisi*, 26(3), 87-96.
- Doğan, N. (2019). *Eğitimde ölçme ve değerlendirme*. Pegem Akademi.
- Doğanay, A. (1997). Ders dinleme sırasında bilişsel farkındalıkla ilgili stratejilerin kullanımı, *Çukurova Üniversitesi Eğitim Fakültesi Dergisi*, 2(15), 34-42.
- Durmuş, Bakıoğlu, B. (2013). *Fen alanı öğretmeni okul ortamı ölçeğinin Türkçeye uyarlanması çalışması*. Yüksek Lisans Tezi, Gaziosmanpaşa Üniversitesi Eğitim Bilimleri Enstitüsü, Tokat.
- Durmuş, Y. T., & Ok, A. (2014). Öğretmenlerin fen ve teknoloji öğretim programına yönelik görüşleri ölçeğinin geliştirilmesi: geçerlik ve güvenirlik çalışması. *Elementary Education Online*, 13(1). Sayfa
- Ekinci, H. Y., & Köksal, E. A. (2011). İlköğretim fen ve matematik öğretmenleri için ölçme ve değerlendirme yeterlikleri ölçeğinin geliştirilmesi. *Kastamonu Eğitim Dergisi*, 19(1), 167-184.
- Erkuş, A. (2016). *Psikolojide ölçme ve ölçek geliştirme*. (3.Baskı). Ankara: Pegem Akademi Yayınları.
- Eroğlu, A. (2008). Faktör analizi. İçinde. Ş. Kalaycı (Ed.), *SPSS Uygulamalı çok değişkenli istatistik teknikleri* (ss. 321-331). Ankara: Asil Yayın Dağıtım.
- Farr, S. (2010). *Teaching as leadership: the highly effective teacher's guide to closing the achievement gap*. San Francisco: John Wiley & Sons.
- Gorsuch, R. L. (1983). *Factor analysis*. Hillsdale: Lawrence Erlbaum Associates.

- Gül, Ş., & Sözbilir, M. (2015). Fen ve matematik eğitimi alanında gerçekleştirilen ölçek geliştirme araştırmalarına yönelik tematik içerik analizi. *Eğitim ve Bilim*, 40(178),85-102.
- Güneş, F. (2012). Öğrencilerin düşünme becerilerini geliştirme. *Türklük Bilimi Araştırmaları*, (32), 127-146.
- Hovardaoğlu, S. (2000). *Davranış bilimleri için araştırma teknikleri*. Ankara: Ve-Ga Yayınevi.
- Kahraman, M., & Polat, D. Fen bilimleri öğretmenlerinin laboratuvar ve öğretim sürecine ilişkin kaygı ölçeği: geçerlik ve güvenirlik çalışması. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, 30(2), 757-780.
- Karasar, N. (2000). *Bilimsel araştırma yöntemleri*. Ankara: Nobel Yayın Dağıtım.
- Kline, P. (1994). *An easy guide to factor analysis*. London and New York: Routledge.
- Kline, R.B., (2005). *Principles and practice of structural equation modeling*, 2nd ed, New York: Guilford Press.
- Koray, Ö., Altunçekiç, A., & Yaman, S. (2005). Fen bilgisi öğretmen adaylarının soru sorma becerilerinin bloom taksonomisine göre değerlendirilmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 17(17), 33-39.
- Krathwohl, D.R., Bloom, B.S., & Masia, B.B. (1964). *Taxonomy of educational objectives: The classification of educational goals. Handbook II: The affective domain*. New York: David McKay.
- Kuzu, B. S. (2016). Öğretmenlerin ölçe değ erlendirme sürecine yönelik tutumlarını ölç en bir ölçek geliştirme çalışması. Yüksek Lisans Tezi, Marmara Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul.
- Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into practice*, 32(3), 131-137.
- McDonald, R. P., & Ho, M. H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64.
- Millî Eğitim Bakanlığı (MEB,2018). İlköğretim kurumları fen bilimleri dersi öğretim programı, Talim ve Terbiye Kurulu Başkanlığı, Ankara.
- Millî Eğitim Bakanlığı (MEB,2021). Geliştirilen ve güncellenen standart kriterlere uygun olarak hazırlanan örnek hizmetiçi eğitim programları. 05 Ocak 2021 tarihinde <http://oygm.meb.gov.tr/dosyalar/StPrg/> adresinden erişildi.
- Moon, J. A. (2004). *A handbook of reflective and experiential learning theory and practice*. London: Routledge Flamer.
- Mutlu, M., Uşak, M., & Aydoğdu, M. (2003). Fen bilgisi sorularının bloom taksonomisine göre değerlendirilmesi. *Kırşehir Eğitim Fakültesi Dergisi*, 4(2), 87-95.
- Özcan, S., & Akcan, K. (2010). Fen bilgisi öğretmen adaylarını hazırladığı soruların içerik ve bloom taksonomisi'ne uygunluk yönünden incelenmesi. *Kastamonu Eğitim Dergisi*, 18(1), 323-330.

- Park, S., & Oliver, J. S. (2008). Revisiting the conceptualisation of pedagogical content knowledge (PCK): PCK as a conceptual tool to understand teachers as professionals. *Research in Science Education*, 38(3), 261–284.
- Pohlmann, J. T. (2004). Use and interpretation of factor analysis in the journal of educational research: 1992-2002. *The Journal of Educational Research*, 98(1), 14-23
- Scherer, R. F., Wiebe F. A., Luther, D. C. & Adams J. S. (1988). Dimensionality of coping: factor stability using the ways of coping questionnaire, psychological reports. *Psychological Reports*, 62(3), 763-770. PubMed PMID: 3406294.
- Shah, J. Y., & Kruglanski, A. W. (2000). Aspects of goal networks: Implications for self-regulation. In. Boekaerts, M., Zeidner, M., & Pintrich, P. R. (Ed.), *Handbook of self-regulation*. Academic Press.
- Simpson, B.J. (1966). The classification of educational objectives: Psychomotor domain. *Illinois Journal of Home Economics*, 10(4), 110-144.
- Sosniak, L. A. (1994). The taxonomy, curriculum, and their relations. *Bloom's taxonomy: A forty-year retrospective*, 103-125.
- Sönmez, V. (1993). *Program geliştirmede öğretmen el kitabı*. Ankara: Adım Yayıncılık.
- Sümer, N. (2000). Yapısal eşitlik modelleri: Temel kavramlar ve örnek uygulamalar. *Türk Psikoloji Yazıları*. 3(6). 49-74.
- Şahin, M.G., & Boztunç Öztürk, N. (2018). Eğitim alanında ölçek geliştirme süreci: Bir içerik analizi çalışması. *Kastamonu Eğitim Dergisi*, 26(1), 191-199. doi:10.24106/kefdergi.375863
- Şahinel, S. (2002), *Eleştirel düşünme*. Ankara: Pegem Yayıncılık.
- Şimşek, Ö.F., (2007). *Yapısal eşitlik modellemesine giriş*. Ankara: Ekinoks.
- Taşkın, G., & Aksoy, G. (2018). Ortaöğretime geçiş sistemi ile ilgili “fen bilimleri öğretmeni görüş ölçeği” geliştirme çalışması. *Eğitim Kuram ve Uygulama Araştırmaları Dergisi*, 4(1), 27-41.
- Tosun, C. & Taşkesenligil, Y. (2015). The instruments used in science education in Turkey: a descriptive content analysis. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 15(2), 364- 383.
- Turgut, M. F., & Baykul, Y. (2012). *Eğitimde ölçme ve değerlendirme*. Pegem Akademi.
- Woolfolk, A. E., Winne, P. H., Perry, N. E., & Shapka, J. (2009). *Educational psychology* (4th Canadian ed.). Upper Saddle River, NJ: Pearson Education.
- Yüksek Öğretim Kurumu (YÖK,2020). Fen bilgisi öğretmenliği lisans programı. 05 Ocak 2021 tarihinde https://www.yok.gov.tr/Documents/Kurumsal/egitim_ogretim_dairesi/Yeni-Ogretmen-Yetistirme-Lisans-Programlari/Fen_Bilgisi_Ogretmenligi_Lisans_Programi.pdf adresinden erişildi.

Ek:Fen Bilimleri Öğretmenlerinin Üst Düzey Öğrenme Düzeylerini Ölçen Soru Geliştirme
Öz-yeterlikleri Ölçeği

Madde No	Maddeler	Yanıtlar				
		1	2	3	4	5
		Kesinlikle	Katılmıyorum	Kararsızım	Katılıyorum	Kesinlikle Katılıyorum
M1	Öğrencilerin problemin amacına uygun olarak gözlem ve ölçüm sonucunda elde ettikleri verileri yazılı ifade, resim, tablo ve çizim gibi çeşitli yöntemlerle kaydedebilme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M2	Öğrencilerin nesneler veya olaylar arasındaki belirgin benzerlikleri ve farklılıkları belirleme becerilerini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M3	Öğrencilerin iki farklı genellemeyi karşılaştırabilme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M4	Öğrencilerin topladığı verilerden sonuç çıkarma düzeylerini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M5	Öğrencilerin elde ettiği bilgilerinden yola çıkarak yeni bir sentezleme (genellemeye ulaşma) becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M6	Öğrencilerin verilen bir olaydaki kontrol değişkenini belirleme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M7	Öğrencilerin elde ettiği ürünü sözlü, yazılı ya da görsel malzeme kullanarak uygun şekillerde sunabilme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M8	Öğrencilerin işledikleri verileri ve oluşturdukları modeli yorumlama becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M9	Öğrencilerin bir çeşit büyüklüğü ölçebilme becerisini belirleyen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M10	Öğrencilerin verilen bir olaydaki bağımsız değişkeni belirleme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M11	Öğrencilerin bir problemin çözümüne yönelik öneri sunabilme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M12	Öğrencilerin karşılaştıkları bir problemin olası çözümünü tahmin edebilme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M13	Öğrencilerin elde ettiği bilgileri birleştirme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M14	Öğrencilerin verilen bir olay veya ilişkide en belirgin bir ya da birkaç değişkeni belirleme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M15	Öğrencilerin gerçekleşmiş olayların sebepleri hakkında gözlemlere dayanarak açıklamalar yapabilme becerilerini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M16	Öğrencilerin kurduğu hipotezi test etme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M17	Öğrencilerin gözlem, çıkarım veya uygulanan deneylere dayanarak geleceğe yönelik muhtemel sonuçlar hakkında tahmin yapabilme becerilerini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M18	Öğrencilerin mevcut bilgileri arasında ilişki kurabilme düzeyini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)

M19	Öğrencilerin farklı kaynaklardan bilgi toplama becerilerini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M20	Öğrencilerin hipotez kurma becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M21	Öğrencilerin yaptığı gözlemlerden yola çıkarak bir veya birden fazla özelliğe göre karşılaştırma yapabilme becerilerini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M22	Öğrencilerin elde ettiği ürünü paylaşabilme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M23	Öğrencilerin benzerlik ve farklılıklara göre grup ve alt-gruplara ayırma becerilerini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M24	Analiz düzeyinde öğrenme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M25	Sentez düzeyinde öğrenme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M26	Öğrencilerin topladığı veriler arasında ilişki kurabilme düzeyini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M27	Öğrencilerin nesneleri sınıflandırmada kullanılacak nitel ve nicel özellikleri belirleme becerilerini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M28	Öğrencilerin olay ve nesnelere yönelik kütle, uzunluk, zaman, sıcaklık ve adet gibi nicelikler için uygun birimleri de belirterek yaklaşık değerler hakkında tahmin etme becerilerini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M29	Öğrencilerin grafik çizimiyle ilgili kuralları uygulama becerilerini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)
M30	Öğrencilerin elde ettiği bilgileri organize etme becerisini ölçen soru maddesi yazabilirim.	(1)	(2)	(3)	(4)	(5)



T.C.
AMASYA ÜNİVERSİTESİ
Bilim Etik Kurulu
Sosyal Bilimler Etik Kurulu



Sayı : 30640013-044
Konu : Etik Onay

Sayın Prof.Dr. Sevilay KARAMUSTAFAOĞLU
Öğretim Üyesi

İlgi : 12/05/2020 tarihli ve 285 sayılı yazı.

"Fen Bilimleri Öğretmenlerinin Üst Düzey Öğrenme Becerilerini Ölçen Soru Yazma Öz Yeterlilik Ölçeği: Geçerlik ve Güvenirlilik" adlı çalışma Sosyal Bilimler Etik Kurul tarafından bilimsel araştırma etiği yönünden değerlendirildi ve incelendi. Konu ile ilgili kurul görüşü ektedir.

Bilgilerinizi rica ederim

e-imzalıdır

Prof.Dr. Halil APAYDIN
Etik Kurul Başkanı

Ek:

- 1- sosyaL-bilimler-etik-kurul-degerlendirme-formu Sevilay Karamustafaoğlu 15 Mayıs (1 sayfa)
- 2- Etik Onay_Sevilay Karamustafaoğlu 15 Mayıs (9 sayfa)

	AMASYA ÜNİVERSİTESİ SOSYAL BİLİMLER ETİK KURUL DEĞERLENDİRME FORMU

Araştırmanın Başlığı: Fen Bilimleri Öğretmenlerinin Üst Düzey Öğrenme Becerilerini Ölçen Soru Yazma Öz Yeterlilik Ölçeği: Geçerlik ve Güvenirlik	
Başvuru Formunun Etik Kurula geldiği tarih	12.05.2020
Başvuru Formunun Etik Kurulda incelendiği tarih	15.05.2020
Karar tarihi	15.05.2020

SONUÇ

1.	☒ Kabul
2.	☐ Düzeltme gereklidir: Etik sorun olabilecek sorular/maddeler, süreçler ya da unsurlar bulunmaktadır. Açıklama:
3.	☐ Red Gerekçe, Görüş, Tavsiye ve Açıklamalar:

Başvuru dosyasının incelenmesinde hazır bulunan ve araştırmayla doğrudan veya dolaylı olarak ilişkisi bulunmayan Etik Kurul başkan ve üyelerinin ad,soyad ve imzaları.

(Başkan)
 Prof.Dr. Halil APAYDIN
 İMZA

(Üye, Bşk. Yrd.)
 Dr. Öğr. Üyesi Burcu KARAŞAR
 İMZA

(Üye)
 Doç.Dr. Songül KEÇECİ KURT
 İMZA

(Üye)
 Doç.Dr. Kemal BAYTEMİR
 İMZA

(Üye)
 Dr. Öğr. Üyesi Kürşat EFE
 İMZA