



Fen Bilimleri Öğretmenleri ve Öğretmen Adaylarına Yönelik Akıl Yürütme Becerileri Testi'nin Geliştirilmesi¹

Development of Scientific Reasoning Skills Test Towards In-service and Pre-service Science Teachers

Merve KOCAGÜL SAĞLAM, Pamukkale Üniversitesi, mervekocagl@gmail.com
Gül ÜNAL ÇOBAN, Dokuz Eylül Üniversitesi, gulunalcoban@gmail.com

Öz. Bu çalışmanın amacı, fen bilimleri öğretmenlerinin ve öğretmen adaylarının fen bilimleri dersinde en çok kullanılan akıl yürütme becerileri konusunda düzeylerinin belirlenmesi için çoktan seçmeli bir test geliştirmektir. Çalışmada, nicel araştırma yöntemlerinden tarama modeli benimsenmiştir. Testin geliştirilme sürecinde, alan yazında var olan akıl yürütme becerilerine yönelik hazırlanmış diğer çoktan seçmeli testler ve çeşitli kaynaklarda yer alan akıl yürütme becerilerini belirleme amacındaki sorular incelenmiştir. 37 adet sorudan oluşan pilot soru formunun kapsam geçerliği uzman görüşü yoluyla sağlanmıştır. Uzman görüşleri doğrultusunda gerçekleştirilen revizyonlardan sonra pilot soru formu üç farklı devlet üniversitesinin fen bilgisi öğretmenliği bölümünde okuyan 432 öğretmen adayına ve çeşitli okullarda görev yapmakta olan 48 fen bilimleri öğretmenine uygulanmıştır. Elde edilen veriler daha sonra sırasıyla madde analizi ve testin yapı geçerliği için tetrakorik korelasyona dayalı açıklayıcı faktör analizine tabi tutulmuştur. Madde analizi sonucu ayırtedicilik indeksi 0,30'dan düşük olan on soru testten çıkarılmıştır. Son hali 27 maddeden oluşan testin orta düzeyde güçlüğü ($p=0,523$) ve yüksek düzeyde ayırtediciliğe ($r_{jx} = 0,480$) sahip olduğu belirlenmiştir. Testin KR-20 güvenilirlik katsayısı ise .812 olarak bulunmuştur. Testin yapı geçerliği sonucunda ise faktör yükleri .24 ile .77 arasında değişen maddelerin tek bir faktör altında toplandıkları belirlenmiştir. Test geliştirme sürecinde elde edilen bulgular geliştirilen Akıl Yürütme Becerileri Testi'nin geçerli ve güvenilir bir ölçme aracı olarak kullanılabileceğini göstermektedir.

Anahtar Sözcükler: Akıl Yürütme Becerileri, Fen Bilimleri Öğretmenleri, Fen Bilgisi Öğretmen Adayları, Test Geliştirme

Abstract. The aim of this study is to develop a multiple choice test for determining inservice and preservice science teachers' scientific reasoning skill levels which are frequently used in science lessons. This study is in survey model which is one of quantitative research methods. During test development process, existing multiple choice tests about scientific reasoning skills and sample questions were examined. Content validity of prepared pilot question form which consisted of 37 items at the beginning was provided via expert opinions. After completing the revisions, pilot form was implemented to 481 preservice science teachers studying at three different state universities and 48 in service science teachers working at various middle schools. After that, obtained data were subjected to item analysis and tetrachoric correlation based on exploratory factor analysis for construct validity respectively. Ten questions which have lower discrimination indexes than .30 were removed from the test at the end of item analysis. The final form of the test which is moderately difficult ($p=0,523$) and has high level of discrimination ($r_{jx} = 0,480$) consisted of 27 items. KR-20 reliability coefficient of the test was found as .812. As a result of construct validity, it was determined that all items whose factor loadings change between .24 and .77 are under one factor. The findings of test development process showed that developed Scientific Reasoning Skills Test can be used as a valid and reliable measurement tool.

Keywords: Scientific Reasoning Skills, Science Teachers, Preservice Science Teachers, Test Development

¹ Bu çalışma, birinci yazarın ikinci yazar danışmanlığında yürüttüğü doktora tezinden üretilmiştir.

SUMMARY

Introduction

Today's societies need individuals having 21st century skills (inquiring, making decision, thinking scientifically, solving real-life problems, reaching scientific knowledge individually and so on). In this context, researches about scientific reasoning skills seem to be very important to gain these skills to individuals. In literature, there are lots of researches about the definition and importance of scientific reasoning skills (Hogan and Fisherkeller, 2005; Kuhn, 2004; Kuhn and Franklin, 2006; NRC, 2012; Osborne, 2013; Sadler and Zeidler, 2004; Schen 2007), students and also teachers' inadequacies (Croker, 2012; Hogan and Maglienti, 2001; Klaczynski, 2000; Klahr and Dunbar, 1988; Nawawra, 2012; Vass, Schiller and Nappi, 2000) about these skills. However, scientific reasoning skills can be developed and improved (Benford and Lawson, 2001; Nashwan, 2005; Vass et al., 2000). So, teachers especially science teachers have a vital role for developing and improving these skills due to the importance of science lessons to gain these skills to students. For this reason, science teachers' scientific reasoning skills knowledge levels need to be determined firstly.

In this context, the aim of this study is to develop a multiple choice test for determining in-service and pre-service science teachers' scientific reasoning skills knowledge levels. When examined the literature, many multiple choice test which intended to measure scientific reasoning skills were found although they had various titles and it was seen that existing multiple choice tests are appropriate for little students, college students or generally adults. Further, it was also found that the content of existing multiple choice tests are different from this developed test. Because the content of this test consists of scientific reasoning skills which are commonly used (inductive reasoning, deductive reasoning, conditional reasoning, correlational reasoning, proportional reasoning, analogical reasoning and causal reasoning) in the context of science lessons.

Method

Survey method which is a quantitative research methods was used in this study because this study aimed to develop Scientific Reasoning Skills Test (SRST) according to test development process. Determination of participants was based on their volunteerism and accessibility, so convenience sampling method was preferred. 432 preservice science teachers studying at three different state universities and 48 science teachers working at either privacy or state middle schools were participated in the study. Pilot form of SRST which includes 37 items was implemented to participants. Data obtained from SRST were scored by authors as 0 for wrong and unanswered items and 1 for correct answers first then analyzed by using Test Analysis Program (TAP) for item analysis and FACTOR program for construct validity.

Results

In the context of validity studies, content and construct validities of test were examined. The content validity of pilot SRST was provided by expert view. Accordingly, it was found that pilot form of SRST consisted of five inductive reasoning items, five deductive reasoning items, five causal reasoning items, five proportional reasoning items, five analogical reasoning items, five conditional reasoning items and seven correlational reasoning items. In the result of item analysis, six items which were low discrimination indexes and six items which had a tendency to increase reliability coefficient were removed from SRST. As a result, it was found that the final form of SRST has 27 items which are moderately difficult and have high level of discrimination. Also it was determined that the final form of SRST includes five items for causal reasoning, four items for inductive reasoning, three items for deductive reasoning, seven items for correlational

reasoning, two items for analogical reasoning, two items for proportional reasoning and four items for conditional reasoning. For construct validity, tetrachoric correlation based exploratory factor analysis was conducted and it was found that all of 27 items whose factor loadings change between .24 and .77 are under one factor. In the context of reliability studies, KR-20 reliability coefficient was calculated and found as .81.

Discussion and Conclusion

In the consequence of analysis, a valid and reliable multiple choice test for measuring scientific reasoning skills knowledge levels was developed. When examined literature, it was seen that this test development process is in accordance with other test development studies generally (Alkan Dilbaz, Özgelen and Yanpar Yelken, 2015; Bingöl and Halisdemir, 2017; Demir, Kızılay and Bektaş, 2016). Many test development studies like this study prefer expert view for providing content validity (Ağgöl Yalçın and Bayrakçeken, 2010; Demir, Kızılay and Bektaş, 2016; Ennis, 1965; Fidan, 2013; Gönen, Kocakaya and Kocakaya, 2011; Hanson, 2016; Tosun and Taşkesengil, 2011) and using tetrachoric correlation based exploratory factor analysis (Baran, 2018; Kutlu ve Karakaya, 2007; Sandal, 2015; Şekercioğlu, Bayat ve Bakır, 2014) for construct validity. However, there were some studies that testing the difference between high and low groups' scores (Akbulut and Çepni, 2013; Ennis, 1965; Gönen, Kocakaya and Kocakaya, 2011; Güven and Topbaş, 2014; Karlı and Ayas, 2013; Ülper, Çetinkaya and Bayat, 2017) and expert view (Alkan Dilbaz, Özgelen and Yanpar Yelken, 2015) for providing construct validity. When examined test development studies in terms of reliability coefficient, many studies in which reports KR-21, Cronbach alpha and correlation coefficients (Ayvaci and Durmuş, 2016; Sundre, 2008; Ülper, Çetinkaya and Bayat, 2017) were found in addition to studies that report KR-20 value (Bingöl and Halisdemir, 2017; Hanson, 2016; Özmen, 2016).

As a consequence, it was thought that SRST is candidate to fill the gap in the context of inservice and preservice science teachers' scientific reasoning skills using levels studies because SRST is both different within scientific reasoning skills types from existing multiple choice test and also is valid and highly reliable measurement tool.

GİRİŞ

Dünyada ve ülkemizde eğitim programları gerçek yaşam problemleriyle baş edebilen, sorgulama, tartışma, karar verme becerisine sahip, bilimsel düşünebilen, bilgiye kendisi ulaşabilen bireylerin topluma kazandırılmasını hedeflemektedir. Bir başka ifadeyle, günümüz toplumları 21. yy becerilerine sahip bireylere ihtiyaç duymaktadır. Bireylere bu becerilerin kazandırılmasında akıl yürütme becerilerine yönelik araştırmalar (Koslowski, 2012) ilgi toplamaktadır.

Akl yürütme becerileri kavramına yönelik, üzerinde tüm araştırmacıların uzlaştığı ortak bir tanım olmamasına rağmen alan yazında çeşitli araştırmacılar tarafından tanımlamalarına rastlanmaktadır. Örneğin; Hogan ve Fisherkeller (2005) akıl yürütmeyi bilimsel bilgi ile ve bilimsel bilgi hakkında düşünme uygulamaları olarak tanımlamışlardır. Bir başka araştırmada akıl yürütme teori ve hipotezler üretmek, araştırmak ve revize etmek için bilginin elde edilmesi ve değiştirilmesi durumu şeklinde tanımlanmıştır (Kuhn ve Franklin, 2006; Wilkening ve Sodian, 2005). Schen (2007) ise akıl yürütmeyi kanıtların toplandığı, analiz edildiği ve kavramlar ve teoriler arasında bağlantıların oluşturulduğu bilimsel bilgi üretiminin kalbi olan bir yöntem olarak tanımlamıştır. Tüm bu tanımlarda ortak olarak bilgi üretimi, revizyonu, bilginin teoriyle koordinasyonu ön plana çıkarılmakta olup aslında örtük olarak bilimsel sorgulama sürecine vurgu yapılmaktadır. Bu sebeple akıl yürütme ile ilgili değişkenlerin kontrolü, tümevarımsal, tümdengelsel, nedensel, ilişkisel, oransal, olasılıksal düşünme becerilerini kapsayan ve bilimsel sorgulamayı destekleyen beceriler şeklinde tanımlar da alan yazında sıklıkla yer almaktadır (Bao vd., 2009; Han, 2013; Zimmerman, 2007).

Akl yürütme becerileri yaparak-yaşayarak bilim öğrenmede merkezi bir öneme sahiptir. Çünkü yaparak-yaşayarak bilim öğrenme, soru sorma, modelleri kullanma ve geliştirme, araştırma planlama, verileri yorumlama ve analiz etme, bilgiyi elde etme ve değerlendirme gibi kanıt tabanlı bilgi iddiaları ve çözümlerini yapılandırma bir dizi bilim ve mühendislik uygulamalarını kapsar (National Research Council [NRC], 2012). Bununla birlikte bu beceriler, bireyleri yaşamlarında karşılaşılabilecekleri sosyo-bilimsel problemleri çözmeye ve iyi bir vatandaş olmaya hazırlar (Osborne, 2013), karar verme süreçlerinde yardımcı olur. Örneğin gideceğimiz yolun mesafesine göre ne kadar yakıt harcayacağımızı hesaplamada orantısal akıl yürütme becerisi, hava durumu tahmininde bulunurken olasılıksal akıl yürütme becerisini kullanırız. Bu sebeple akıl yürütme becerisinin bilimsel okuryazar birey profili oluşturulmasına katkısı yadsınamaz öneme sahiptir. Günlük yaşamda sağladığı yararların yanı sıra akıl yürütme becerileri akademik anlamda da önemli bir yere sahiptir. Çünkü bu beceriler kavramsal anlamının kalitesini yansıtır (Lawson, 2005; Sadler ve Zeidler, 2004), akademik başarının (Coletta ve Philips, 2005), bilgi edinimi ve kavramsal değişim sürecinin (Kuhn, 2004) belirleyicisi olarak rol oynar. Chen ve She (2015) çalışmalarında açık bir şekilde akıl yürütme eğitimi alan 5. Sınıf öğrencilerinin bu eğitimi almayan öğrencilere göre anlamlı derecede daha test edilebilir ve doğru hipotezler, kanıt tabanlı açıklamalar ürettikleri ve akıl yürütme seviyelerinin arttığını bulmuşlardır.

Bu denli öneme sahip olmasına rağmen pek çok araştırma öğrencilerin akıl yürütme becerileri konusunda yetersizliğini konu edinmektedir (Crocker, 2012; Hogan ve Maglienti, 2001; Klaczynski, 2000; Vass, Schiller ve Nappi, 2000). Öğrencilerin akıl yürütme konusunda sahip oldukları bu yetersizlik ve eksikliklerin kaynağı olarak ise en fazla teori-kanıt koordinasyonunu sağlayamamaları (Kuhn, 2004), inandıklarını kesin doğrular olarak kabul etmeleri (Klahr ve Dunbar, 1988) gösterilmektedir. Bununla birlikte akıl yürütme becerileri sonradan geliştirilebilen becerilerdir. Nitekim Schwartz, Lederman ve Crawford (2004) öğretmen adaylarıyla gerçekleştirdikleri çalışmalarında akıl yürütme ile ilgili doğrudan fırsatlar sunulmasının ve onlara düşünme için yeterli zaman tanıma ve tartışma imkânı sunma gibi öğretmenin rehber davranışlarının öğretmen adaylarının akıl yürütme ile ilgili görüşlerini güçlendirdiğini rapor etmişlerdir. Benzer şekilde Vass vd. (2000) etkili bir fen öğretiminin öğrencilerin orantısal, olasılıksal ve ilişkisel akıl yürütme becerilerindeki eksiklikleri iyileştirebildiğini bildirmişlerdir. Nashwan (2005) ise akıl yürütme becerilerini geliştirmeye yardımcı öğretimsel deneyimler olarak merak uyandırma, keşif ve planlamayı belirtmiştir.

Şüphesiz ki etkili bir öğretim, iyi bir öğretici ile mümkün olur. Bu sebeple akıl yürütme becerilerini geliştirici öğretimde öğretmen rolü göz ardı edilemez. Çünkü öğretmenin öğrencilerden gelecek farklı sorulara gösterdiği tolerans, öğrenci sorularına verdiği yetersiz yanıtlar, öğrencilerden koşulsuz bir şekilde söylenenlere itaat etmelerini beklemek, öğretmenin sunduğu değerlendirme yöntemi (çoktan seçmeli test vb.), öğrencileri aktif öğrenme ortamlarına dâhil etme ya da etmeme gibi sergilediği davranış ve tutumlar da öğrencilerin akıl yürütme becerileri gelişimi üzerinde önemli bir pay sahibidir. Ayrıca öğretmenin sahip olduğu akıl yürütme becerisi seviyesi de öğrencilerin akıl yürütme becerisi gelişimleri açısından önemli bir etkidir. Benford ve Lawson (2001) daha yüksek düzey akıl yürütme becerisine sahip öğretmenlerin öğrencilerine sorgulamaya dayalı daha etkili öğrenme ortamları sunabildiğini bildirmiştir. Diğer yandan öğretmenlerin istenilen düzeyde bilimsel düşünme becerisine sahip olmadığını belirten çalışmalar da bulunmaktadır (Nawawra, 2012).

Akıl yürütme becerileri sonradan geliştirilebilen becerilerdir ve gerek akademik anlamda gerek günlük yaşamda önemli bir yere sahiptir. Öğrencilere bu becerilerin kazandırılmasında öğretmen büyük bir role sahiptir. Buradan hareketle akıl yürütme becerilerinin kazandırılabilmesi için temel derslerden birisi olan fen bilimleri dersi öğretmenlerinin akıl yürütme becerileri düzeylerinin belirlenmesi oldukça önemlidir. Alan yazın incelendiğinde farklı isimlerle yer alsalar da akıl yürütme becerilerini ölçen çoktan seçmeli testler olduğu görülmüştür. Bu testlerden birisi akıl yürütme becerileri ile ilgili araştırmalarda sıklıkla referans verilen Lawson Classroom Test of Scientific Reasoning'dir. Bu test Lawson (1978) tarafından öğrencilere yönelik olarak geliştirilmiş, Ateş (2002) tarafından dilimize uyarlanmıştır. Test kombinasyonel düşünme, değişkenlerin belirlenmesi ve kontrolü, olasılıksal düşünme, korelasyonel düşünme ve oranlı düşünme olarak beş zihinsel yetenek ile korunum yasasını kavramayı ölçen sorulardan oluşmaktadır. Sorularda şekil veya sözlü olarak anlatılan bir duruma yönelik cevabın istendiği çoktan seçmeli cevaplar ve daha sonra verilen cevabın yine çoktan seçmeli olarak istendiği gerekçe bölümü bulunmaktadır. Bir başka test Watson ve Glaser (1964) tarafından dokuzuncu sınıftan yetişkinliğe kadar olan kişilerin eleştirel düşünme becerilerini ölçme amaçlı geliştirilmiş olan Watson-Glaser Eleştirel Akıl Yürütme Gücü Testi'dir. Çoktan seçmeli olan bu test çıkarsama, varsayımları farkına varma, tümevarım, yorumlama ve karşıt görüşleri değerlendirmeye yönelik sorulardan oluşmaktadır. Facione (1990) tarafından geliştirilen Kalifornia Eleştirel Düşünme Becerileri Testi ise üniversite öğrencilerine yönelik hazırlanmış yorum yapma, denenceler ve tahminleri analiz etme ve tümevarım becerilerini ölçen çoktan seçmeli sorulardan oluşmaktadır. Bu alanda geliştirilmiş bir diğer test Cornell Eleştirel Düşünme Testi'dir (Ennis, Millman ve Tomko, 1985). Bu testin 4-14. sınıf seviyesine uygun X ve üniversite öğrencileriyle yetişkinlere uygun Z olarak iki seviyesi mevcuttur. Her iki formda çoktan seçmeli sorulardan oluşmakla birlikte X seviyesi tümevarım, güvenilirlik, gözlem, tümdengelim ve varsayımın tanımlanmasına yönelik soruları; Z seviyesi ise X seviyesindeki sorulara ilaveten tahmin, deneysel planlama ve yanılgılara yönelik sorular da içermektedir. New Jersey Akıl Yürütme Becerileri Testi ise 4. sınıftan üniversiteye kadar olan öğrenciler için tasarlanmıştır (Hecht, 1980). Çoktan seçmeli olan bu test tümevarım, varsayımları tanımlama ve gerekçelendirme sorularını içermektedir. Tobin ve Capie (1981) tarafından geliştirilmiş ve Geban, Aşkar ve Özkan (1992) tarafından dilimize uyarlanmış olan Mantıksal Düşünme Testi (Test of Logical Thinking) değişkenlerin kontrolü, orantısal akıl yürütme, olasılıksal akıl yürütme, ilişkisel akıl yürütme ve kombinasyonel akıl yürütmeyi ölçen toplam 10 sorudan oluşmaktadır. Bir diğer test olan Inquiry for Scientific Thinking and Reasoning Assessment (iSTAR) (2010) üçüncü sınıftan üniversite 1. sınıfa kadar olan kişilerin akıl yürütme becerilerini ölçme amaçlı çoktan seçmeli yaklaşık 300 sorudan oluşmaktadır. Bu test kapsamında tümdengelimsel akıl yürütme, nedensel akıl yürütme, koşulsal akıl yürütme, değişkenlerin kontrolü, ilişkisel akıl yürütme, gizli değişkenler, hipotetik-tümdengelimsel akıl yürütme, tümevarımsal akıl yürütme, olasılık ve oran-orantı konularına yönelik sorular yer almaktadır. Sundre (2008) tarafından geliştirilen Madison Assessment: The Scientific Reasoning Test Version 9, 49 soruluk çoktan seçmeli bir testtir. Üniversite öğrencilerinin genel akıl yürütme becerilerini ölçmek amacıyla geliştirilmiştir. Sorular herhangi bir konuya özel değildir. Bilimi sözde bilimden ayırma, tahminde bulunma ve tahminleri anlamak için teori ve modelleri kullanma, uygulamalı araştırmaların ve teknolojinin birbirine

Şekil 1. Test geliştirme süreci

Testin amacının belirlenmesi: Akıl Yürütme Becerileri Testi (AYBT)'nin amacı, fen bilimleri öğretmenlerinin ve fen bilgisi öğretmen adaylarının akıl yürütme becerilerini kullanma düzeylerini belirlemektir.

Ölçülecek özelliklerin belirlenmesi: AYBT kapsamında fen bilimleri dersinde en çok kullanılan akıl yürütme becerileri olarak tümevarımsal, tümdengelimsel, nedensel, ilişkisel, orantısal, koşulsal ve analogik akıl yürütme becerilerinin ölçülmesine yönelik sorular hazırlanmıştır. Alan yazında söz konusu akıl yürütme becerilerinin tümünü içeren bir ölçme aracına rastlanmamış olması geliştirilen testin alandaki bu boşluğu kapatma yolunda önemini de ortaya koymaktadır.

Maddelerin yazılması: Öncelikle söz konusu akıl yürütme becerilerini ölçmek amacıyla soruların çoktan seçmeli olarak hazırlanmasına karar verilmiştir. Bunun sebebi, çoktan seçmeli soruların cevaplanmasının ve puanlanmasının daha kısa süre alması ve nesnel bir şekilde puanlanmasıdır (Özdamar, 2016). Soru içeriklerinin oluşturulmasında ise alan yazında yer alan akıl yürütme becerileri testlerindeki sorular ve çeşitli kaynaklarda (Lawson, 1978; iSTAR, 2010; URL-1) bulunan örnek sorular incelenmiştir. İncelemelerden sonra birçoğu fen alanındaki konuları, birkaçı ise alan dışı konuları kapsayan 26 farklı konuda -bazı maddeler aynı soru öbeğine bağlı olarak- toplam 37 madde hazırlanmıştır.

Maddelerin gözden geçirilmesi: 37 maddeden oluşan test hazırlandıktan sonra kapsam geçerliğini sağlamak üzere uzman görüşüne başvurulmuştur. Bu kapsamda alanında uzman bir profesör, bir doçent, iki araştırma görevlisi ve bir fen bilimleri öğretmenine hazırlanan 37 maddelik soru formu verilmiştir ve her bir sorunun hedeflenen akıl yürütme becerisini ölçüp ölçmediği üzerine görüş bildirmesi istenmiştir. Uzmanlardan gelen görüşler ve düzeltmeler doğrultusunda bazı soru maddeleri gözden geçirilerek gerekli düzeltmeler gerçekleştirilmiştir.

Deneme formunun hazırlanması: Uzman görüşü sonrası gerekli düzenlemeleri yapılan 37 maddelik soru formuna testin amacının, nerede kullanılacağı ve nasıl cevaplanacağı açıkladığı kısa bir yönerge eklenmiştir. Bu şekliyle test pilot uygulamaya hazır hale getirilmiştir.

Deneme formunun uygulanması: Araştırmacı tarafından fen bilimleri öğretmenleri ve öğretmen adaylarına yönelik olarak hazırlanan test pilot uygulama için fen bilgisi öğretmenliği lisans programında öğrenim görmekte olan 432 fen bilgisi öğretmen adayına ve ayrıca 48 fen bilimleri öğretmenine uygulanmıştır. Test uygulanmadan önce, uygulama grubuna testin amacı ve nasıl yanıtlanacağı ile ilgili kısa bir açıklama yapılmıştır. Testin uygulanması yaklaşık 30-35 dakika sürmüştür.

Testin puanlanması, madde analizi ve seçimi: Test uygulandıktan sonra doğru cevaplara 1 puan, yanlış ve boş cevaplara ise 0 puan verilerek puanlanmıştır. Daha sonra test sırasıyla madde gücü ve ayrılcılığı açısından incelenmek üzere Test Analiz Programında (TAP) (Brooks ve Johanson, 2003) ve sonrasında yapı geçerliğini sağlamak üzere FACTOR programında (Lorenzo-Seva ve Ferrando, 2013) tetrakorik korelasyona dayalı açımlayıcı faktör analizine tabi

tutulmuştur. Yapı geçerliği ve madde analizi kapsamında gerçekleştirilen işlemler “Bulgular” başlığı altında sunulmuştur.

BULGULAR

AYBT'nin Kapsam Geçerliğine İlişkin Bulgular

Kapsam geçerliği, testi oluşturan maddelerin ölçmeyi hedeflediği özellikleri ölçme konusunda nitelik ve nicelik olarak yeterliğini ifade eden bir kavram olarak ifade edilmektedir (Büyüköztürk, 2012). AYBT'nin kapsam geçerliği uzman görüşü yoluyla sağlanmıştır. Bu amaçla hazırlanan 37 maddelik soru formu fen bilgisi eğitimi alanında uzman bir profesör, bir doçent, iki araştırma görevlisi ve bir fen bilgisi öğretmeninin görüşüne sunulmuştur ve kendilerinden hazırlanan soruların hedeflenen akıl yürütme becerisini ölçüp ölçmediği üzerine görüş bildirmeleri ve ayrıca soruları çözmeleri istenmiştir. Gelen görüşler doğrultusunda soruların ölçmeyi hedeflediği becerilere yönelik herhangi bir sorun olmadığı bildirilmiş, bununla birlikte bazı sorular daha kolay anlaşılabilmesi amacıyla gelen öneriler doğrultusunda yeniden düzenlenmiştir. Testte yer alan maddelerin ölçmeyi amaçladığı akıl yürütme becerilerine göre dağılımı Tablo 1’de verilmiştir.

Tablo 1. Test maddelerinin ölçülecek akıl yürütme becerilerine göre dağılımı

Akıl yürütme becerileri	İlgili olduğu test maddeleri
Tümevarımsal akıl yürütme	S2, S9, S13, S18, S29
Tümdengelimsel akıl yürütme	S3, S19, S22, S30, S33
Nedensel akıl yürütme	S11, S17, S23, S24, S35
Orantısal akıl yürütme	S1, S4, S5, S36, S37
Koşulsal akıl yürütme	S7, S12, S15, S20, S31
İlişkisel akıl yürütme	S6, S10, S14, S16, S21, S32, S34
Analojik akıl yürütme	S8, S25, S26, S27, S28

Tablo 1 incelendiğinde AYBT madde havuzunda tümevarımsal, tümdengelimsel, nedensel, orantısal, analojik ve koşulsal akıl yürütme becerisini ölçen beşer, ilişkisel akıl yürütme becerisine yönelik ise yedi madde olduğu görülmektedir.

AYBT'nin Madde Analizine İlişkin Bulgular

AYBT madde analizi, Test Analiz Programı (TAP) kullanılarak gerçekleştirilmiştir. Gerçekleştirilen test istatistikleri sonucu (Tablo 2) ve her bir maddeye ilişkin elde edilen güçlük ve ayırıcılık indeksleri (Tablo 3) aşağıda sunulmuştur.

Tablo 2. Akıl yürütme becerileri testi analiz sonuçları

N	(\bar{X})	(SS)	p_{ort}	r_{jx}	KR-20
481	18,489	5,656	0,500	0,361	0,764

Tablo 3. Testte yer alan her bir maddeye ilişkin güçlük ve ayırıcılık indeksleri

Madde No	p	r_{jx}	Madde No	p	r_{jx}
S1.	0,52	0,35	S14.2	0,36	0,46
S2.	0,52	0,16	S15.1	0,62	0,71
S3.	0,69	0,07	S15.2	0,52	0,53
S4.	0,83	0,20	S15.3	0,54	0,47

S5.	0,79	0,26	S16	0,27	0,48
S6.	0,26	0,01	S17	0,39	0,39
S7.	0,70	0,51	S18	0,43	0,38
S8.	0,42	0,12	S19.1	0,19	0,01
S9.	0,57	0,60	S19.2	0,72	0,52
S10.1	0,55	0,45	S20	0,23	-0,01
S10.2	0,39	0,30	S21.1	0,56	0,52
S11.1	0,39	0,36	S21.2	0,23	-0,08
S11.2	0,65	0,62	S21.3	0,42	0,32
S12.1	0,69	0,58	S22.1	0,57	0,42
12.2	0,21	0,09	S22.2	0,56	0,55
S13.1	0,59	0,31	S23.1	0,39	0,39
S13.2	0,54	0,54	S23.2	0,37	0,41
S14.1	0,73	0,47	S24	0,42	0,35
			S25	0,65	0,50

Test geliřtirmede madde seçiminde göz önünde bulundurulması gereken bazı kriterler arařtırmacılar tarafından açıklanmıřtır. Buna göre; testte yer alacak maddelerin ayırıcılık indeksinin 0.19 ve altında olması o maddenin o test içerisinde yer almaması gerektięi; 0.20 ve 0.29 arasında olması o maddenin düzeltilerek test içerisinde yer alabileceęi; 0.30 ve 0.39 arasında olması maddenin ayırıcılıęının iyi düzeyde olduęu ve 0.40 ve üzerinde olması o maddenin bilen ve bilmeyeni çok iyi düzeyde ayırabildięi řeklinde yorumlanır (Tekin, 2003; Turgut ve Baykul, 2010). Ayrıca testte yer alacak maddelerin ortalama güçlüklerinin 0.5 düzeyinde olması istenir (Tekin, 2003; Turgut ve Baykul, 2010). Bir dięer kriter ise testin güvenilirlięi ile ilgilidir. Buna göre çoktan seçmeli testlerde olduęu gibi doęru cevaba 1, yanlıř ve boş cevaplara 0 verilerek puanlama yapılan ölçme araçlarında güvenilirlik katsayısı olarak kullanılan KR-20'nin minimum 0.60 olması (Özdamar, 2016), 0.70'den büyük KR-20 deęerinin yüksek güvenilirlik göstergesi olduęu belirtilmiřtir (Büyüköztürk, 2012; Özdamar, 2016). Tablo 2 incelendięinde elde edilen deęerler testin ortalama güçlükte ve iyi düzeyde ayırt edicilięe sahip olduęunu ve güvenilirlięinin kabul edilebilir olduęunu (KR-20 > 0.70) göstermekle birlikte Tablo 3 incelendięinde bazı maddelerin (2, 3, 4, 5, 6, 8, 12.2, 19.1, 20 ve 21.2 sorular) ayırıcılık indekslerinin 0.30'un olduęu göze çarpmaktadır. Bu maddeler çıkarıldıktan sonra tekrarlanan analizde elde edilen sonuçlar Tablo 4'de sunulmuřtur.

Tablo 4. Ayırıcılık indeksi düşük olan maddeler çıkarıldıktan sonra testin analiz sonuçları

N	(\bar{X})	(SS)	Port	r_{jxort}	KR-20
481	14,119	5,383	0,523	0,480	0,812

Tablo 5. Çıkarılan maddeler sonrası her bir maddeye iliřkin güçlük ve ayırıcılık indeksleri

Madde No	p	r_{jx}	Madde No	p	r_{jx}
S1.	0,52	0,40	S15.2	0,52	0,56
S7	0,70	0,52	S15.3	0,54	0,51

S9	0,57	0,63	S16	0,27	0,44
S10.1	0,55	0,44	S17	0,39	0,37
S10.2	0,39	0,32	S18	0,43	0,44
S11.1	0,39	0,34	S19.2	0,72	0,54
S11.2	0,65	0,67	S21.1	0,56	0,57
S12.1	0,69	0,59	S21.3	0,42	0,37
S13.1	0,59	0,37	S22.1	0,57	0,42
S13.2	0,54	0,53	S22.2	0,56	0,58
S14.1	0,73	0,48	S23.1	0,39	0,42
S14.2	0,36	0,40	S23.2	0,37	0,47
S15.1	0,62	0,77	S24	0,42	0,25
			S25	0,65	0,56

Tablo 5 incelendiğinde maddelerin ayırıcılık indekslerinin 0.30'un üzerinde olduğu görülmekle birlikte yalnızca bir maddenin (24. Madde) ayırıcılık indeksinin kabul edilen değerlerin altında kaldığı görülmüştür. Yapılan incelemede maddenin testten çıkarılmasının kapsam geçerliği açısından uygun olmadığı düşünülmüş ve ayrıca söz konusu madde çıkarımının test güvenilirliğini yükseltmediği görülmüştür. Bu sebeple maddenin testte kalmasına karar verilmiştir.

Bu haliyle geliştirilen testin kolay, orta ve zor maddeleri bir arada barındıran orta güçlükte ve yüksek ayırıcılıkta bir test olduğu saptanmıştır. KR-20 güvenilirlik katsayısı 0.81 olarak hesaplanan 27 maddelik bu testin dört sorusu tümevarımsal akıl yürütme becerisini; üç sorusu tümdengelsel akıl yürütme becerisini; beş sorusu nedensel akıl yürütme becerisini; yedi sorusu ilişkisel akıl yürütme becerisini; iki sorusu analogik akıl yürütme becerisini; iki sorusu orantısal akıl yürütme becerisini ve dört sorusu ise koşulsal akıl yürütme becerisini ölçmeye yöneliktir.

AYBT'nin Yapı Geçerliğine İlişkin Bulgular

Bir ölçme aracının ölçülecek olguyu tam ve yansız olarak ölçmesi ile ilgili bir kavram olarak karşımıza çıkan yapı geçerliğinin sağlanmasıyla ilgili Özdamar (2016) üç farklı yöntem önermiştir. Bunlar;

- Ölçme aracının performansının yeterliliğinin kavramsal olarak değerlendirilmesi,
- Tanımlanmış yapının altında yatan teoriden ölçme aracının performansı ile ilgili sınanabilir hipotezler kurulması ve
- Kurulan hipotezleri sınamak için deneysel ve istatistiksel çalışmaların yapılmasıdır.

Benzer şekilde Atılgan, Kan ve Doğan (2016) da yapı geçerliğinin bir testi oluşturan alt testlerle testin bütünü arasındaki korelasyona bakılması, aynı yapıyı ölçen diğer testlerle korelasyonuna bakılması, testin farklı olduğu bilinen gruplara uygulanması yoluyla elde edilen puanların istatistiksel anlamlı bulunması, test-tekrar test yönteminin uygulanması ve faktör analizi yapılması yoluyla sağlanabileceğini belirtmiştir. Çalışma kapsamında geliştirilen testin yapı geçerliğinin sağlanmasında tetrakorik korelasyona dayalı açılıcı faktör analizi tercih edilmiştir. Analiz işlemi Lorenzo-Seva ve Ferrando (2013) tarafından geliştirilen FACTOR programı kullanılarak yapılmıştır. Açılıcı faktör analizinin tetrakorik korelasyona dayalı gerçekleştirilmesinin sebebi, 0-1 şeklinde puanlanan ölçme araçlarında maddeler arasındaki korelasyon katsayısının belirlenmesinde bu tekniğin tercih edilmesidir (Olsson, 1979). Faktör

analizini gerçekleştirmede ortak faktör sayısını belirlemede Hull yöntemi tercih edilmiştir. Bunun sebebi, bu yöntemin faktör sayısını belirlemede kullanılan diğer yöntemlerden (paralel analiz vb.) daha üstün performans gösterdiği bulgusudur (Lorenzo-Seva, Timmerman ve Kiers, 2011). Faktör çıkarma yöntemi olarak ise Minimum sıra faktör analizi (minimum rank factor analysis, MRFA) seçilmiştir. Bunun sebebi, yöntemin açıklanan ve açıklanamayan varyans oranlarını görmeye imkân tanınması (Socan, 2003) ve bu sayede her bir faktör tarafından açıklanan varyans oranının yorumlanabilir olmasıdır (Lorenzo-Seva ve Ferrando, 2006). Faktör sayısı ve faktör çıkarma yöntemlerine karar verildikten sonra programda faktör döndürme yöntemi olarak test maddelerinin birbirinden tamamen birbirinden bağımsız olmadığı gerekçesiyle eğik döndürme (oblique rotation), teknik olarak Promin ve faktör yükü olarak .30 ve üzeri (Büyüköztürk, 2012) tercih edilmiştir. Bu koşullar altında gerçekleştirilen faktör analizi kapsamında öncelikle verilerin faktör analizine uygunluğu Kaiser-Meyer-Olkin (KMO) katsayısı ve Barlett küresellik test sonuçlarına göre değerlendirilmiş ve KMO katsayısının .87 ve Barlett testi'nin anlamlı olduğu bulunmuştur. Verilerin faktör analizine uygunluğundan söz edebilmek için KMO katsayısının .60'ın üzerinde Barlett testi'nin anlamlı çıkması gerekir (Tabachnik ve Fidell, 2007). Buna göre verilerin faktörleşmeye uygun olduğu söylenebilir. Gerçekleştirilen analiz sonucunda program testin tek boyutlu olmasını önermiştir. Bu haliyle açıklanan varyans oranı %37, 63 olduğundan ve alan yazında tek faktörün açıkladığı varyans oranının minimum %30 olması gerektiği görüşünden (Büyüköztürk, 2012) hareketle testin tek boyutlu olabileceği kanısı güçlenmiştir. Daha sonra ise tek boyutluluğu test etmek üzere ek analiz işlemleri gerçekleştirilmiştir. Gerçekleştirilen ek analizler sonucunda tek boyut uygunluk (unidimensional congruence, UniCo), açıklanan ortak varyans (explained common variance, ECV) ve madde artık mutlak yükleri ortalaması (mean of item residual absolute loadings, IREAL) değerleri incelenmiştir. Alan yazında bir testin tek boyutlu olarak kabul edilebilmesi için hem testin geneli hem de madde bazında UniCo değerinin .95 ve üzerinde; ECV değerinin .70 ve .85 aralığında olması ve IREAL değerinin .30'un altında olması gerektiği belirtilmektedir (Ferrando ve Lorenzo-Seva, 2017). Gerçekleştirilen analiz kapsamında testin geneli için %95 güven aralığında .943 ile .955 aralığında olması gereken UniCo değeri .949; .844 ile .861 aralığında olması gereken ECV değeri .847 ve .15 olması gereken IREAL değeri .15 olarak bulunmuştur. Buradan hareketle testin geneli için tek boyutluluk koşulunun sağlanmış olduğu görülmekle birlikte madde bazında yapılan incelemelerde bazı maddelerin bazı tek boyutluluk değerlendirme kriterlerine uygun olmadığı görülmüştür. Bununla birlikte test maddelerinin büyük çoğunluğunda (27 maddeden 25'i) UniCo, ECV ve IREAL değerleri sağlanmış olduğundan test tek boyutlu olarak kabul edilmiştir. Bu haliyle geliştirilen testin faktör yükleri .24 ile .77 arasında değişmektedir. Burada faktör yükleri .30'un altında olan iki madde (5 ve 26) dikkat çekmekle birlikte yük değerleri %95 güven aralığı kapsamında belirtilen aralığa (5. madde için .14 ile .42; 26. madde için .07 ile .36) uygun olduğundan ve ortak varyans değerleri (communalities) .50'nin üzerinde olduğundan bu maddeler problemliler olarak değerlendirilmemiştir.

TARTIŞMA ve SONUÇ

Fen bilimleri öğretmenlerinin ve fen bilgisi öğretmen adaylarının fen derslerinde sıklıkla kullanılan akıl yürütme becerilerine yönelik bilgi düzeylerini ölçebilecek çoktan seçmeli bir test geliştirmenin amaçlandığı bu çalışmada testin amacının ve ölçülecek özelliklerin belirlenmesi, maddelerin yazılması ve gözden geçirilmesi, deneme formunun hazırlanması ve uygulanması ve madde analizinin gerçekleştirilmesi adımları izlenerek test geliştirilmiştir (Turgut ve Baykul, 2010). Alan yazın incelendiğinde geliştirilen çeşitli çoktan seçmeli ölçme araçlarında da benzer test geliştirme adımlarının izlendiği görülmektedir (Şen ve Eryılmaz, 2011; Karlı ve Ayas, 2013; Alkan Dilbaz, Özgelen ve Yanpar Yelken, 2015; Demir, Kızılay ve Bektaş, 2016; Bingöl ve Halisdemir, 2017).

Bir ölçme aracı geliştirilirken göz önünde bulundurulması gereken iki temel nokta ölçme aracının geçerliğinin ve güvenilirliğinin sağlanmasıdır (Turgut ve Baykul, 2010; Büyüköztürk, 2012; Atılğan, Kan ve Doğan, 2016; Özdamar, 2016). Bu çalışma kapsamında geliştirilen çoktan seçmeli testin geçerliğinin sağlanmasında kapsam ve yapı geçerliği incelenmiş ve güvenilirlik

hesaplamalarında KR-20 katsayısı kullanılmıştır. Geliştirilen testin kapsam geçerliğinin sağlanmasında uzman görüşüne başvurulmuştur. Alan yazında geliştirilen diğer çoktan seçmeli testlerin kapsam geçerliğinin sağlanmasında da uzman görüşüne başvuran çeşitli çalışmalar mevcuttur (Ennis, 1965; Akkuş ve Duatepe Paksu, 2006; Ağgöl Yalçın ve Bayrakçeken, 2010; Gönen, Kocakaya ve Kocakaya, 2011; Tosun ve Taşkesengil, 2011; Fidan, 2013; Karşı ve Ayas, 2013; Demir, Kızılay ve Bektaş, 2016; Hanson, 2016). Bir ölçme aracının yapı geçerliğinin sağlanmasında izlenebilecek yöntemlerden birisi olarak faktör analizi yapılmasıdır (Özdamar, 2016). Bu çalışma kapsamında geliştirilen çoktan seçmeli testin yapı geçerliği, tetrakorik korelasyona dayalı açıklayıcı faktör analizinin yanında ayrıca madde analizi yoluyla da sağlanmıştır.

Özellikle çoktan seçmeli ölçme aracı geliştirilmesini ve incelenmesini içeren çalışmalarda bu çalışmada tercih edilen yapı geçerliğini sağlama yönteminin kullanıldığı görülmektedir (Baran, 2018; Kutlu ve Karakaya, 2007; Sandal, 2015; Şekercioğlu, Bayat ve Bakır, 2014). Bununla birlikte yapı geçerliğinin sağlanmasında üst ve alt grup puanlarının istatistiksel yollarla karşılaştırılması (Ennis, 1965; Gönen, Kocakaya ve Kocakaya, 2011; Akbulut ve Çepni, 2013; Karşı ve Ayas, 2013; Güven ve Topbaş, 2014; Ülper, Çetinkaya ve Bayat, 2017). ve uzman görüşüne başvurma (Alkan Dilbaz, Özgelen ve Yanpar Yelken, 2015) yöntemlerinden de yararlanılmıştır. Geliştirilen testin güvenilirlik hesaplamalarında KR-20 katsayısı kullanılmıştır. KR-20 katsayısı madde seçeneklerinin bu testte olduğu gibi 1-0 şeklinde puanlandığı ve soruların güçlük derecelerinin birbirinden farklı olduğu durumlarda kullanılan bir iç tutarlık katsayısıdır (Atılğan, Kan ve Doğan, 2016; Özdamar, 2016) ve bu çalışma kapsamında geliştirilen testin KR-20 güvenilirlik katsayısı .812 olarak bulunmuştur. Alan yazında yer alan diğer çoktan seçmeli testler incelendiğinde güvenilirlik katsayısı olarak KR-20'nin kullanıldığı çalışmalar (Karşı ve Ayas, 2013; Alkan Dilbaz, Özgelen ve Yanpar Yelken, 2015; Özmen, 2016; Hanson, 2016; Bingöl ve Halisdemir, 2017) bulunmakla birlikte güvenilirlik hesaplamalarında Cronbach alpha katsayısının (Akkuş ve Duatepe Paksu, 2006; Aybek ve Çelik, 2007; Sundre, 2008; Şen ve Eryılmaz, 2011; Demir, 2013; Karşı ve Ayas, 2013; Ayvaci ve Durmuş, 2016; Demir, Kızılay ve Bektaş, 2016; Ülper, Çetinkaya ve Bayat, 2017), Spearman-Brown iki yarı test güvenilirliğinin (Gönen, Kocakaya ve Kocakaya, 2011; Ülper, Çetinkaya ve Bayat, 2017) ve korelasyon katsayılarının da (Ateş, 2002; Akbulut ve Çepni, 2013) kullanıldığı çalışmalara da rastlanmıştır. Ayrıca KR-20 güvenilirlik katsayısını rapor eden çalışmalarda güvenilirlik katsayılarının 0.51 ile 0.85 arasında bulunduğu görülmüştür. Bu açıdan değerlendirildiğinde geliştirilen testin alan yazında elde edilen değerlerden yüksek güvenilirlik katsayısına yakın olduğu görülmektedir (.812).

Sonuç olarak geliştirilen Akıl Yürütme Becerileri Testi'nin hem ölçmeyi amaçladığı akıl yürütme becerisi türleri bakımından alan yazında var olan diğer testlerden farklılaşması hem de geçerli ve güvenilirliği yüksek bir ölçme aracı olması sebebiyle fen bilimleri öğretmenlerinin ve öğretmen adaylarının akıl yürütme becerilerini kullanma düzeylerini konu alacak çalışmalar bağlamında alan yazındaki önemli bir boşluğu doldurmaya aday olacağı düşünülmektedir. Geliştirilen test fen bilimleri öğretmenleri ve öğretmen adaylarının akıl yürütme becerileri bilgi düzeylerinin belirlenmesi amacıyla gerçekleştirilecek çeşitli araştırmalarda kullanılabilir. Bu çalışmalardan elde edilen veriler doğrultusunda öğretmenlerin ve öğretmen adaylarının hangi akıl yürütme becerisinin geliştirilmesi konusunda daha fazla desteğe ihtiyaç duyduğu belirlenebilir, iyileştirme amacıyla çeşitli eğitim programları düzenlenebilir.

KAYNAKÇA

- Ağgöl Yalçın, F. ve Bayrakçeken, S. (2010). 5E öğrenme modelinin fen bilgisi öğretmen adaylarının asit-baz konusu başarılarına etkisi. *International Online Journal of Educational Sciences*, 2(2), 508-531.
- Akbulut, H. İ. ve Çepni, S. (2013). Bir üniteye yönelik başarı testi nasıl geliştirilir?: İlköğretim 7. Sınıf Kuvvet ve Hareket Ünitesi. *Amasya Üniversitesi Eğitim Fakültesi Dergisi*, 2 (1), 18-44.
- Akkuş, O. ve Duatepe Paksu, A. (2006). Orantısız akıl yürütme becerisi testi ve teste yönelik dereceli puanlama anahtarı geliştirilmesi. *Eurasian Journal of Educational Research*, 25, 1-10.
- Alkan Dilbaz, G., Özgelen, S. ve Yanpar Yelken, T. (2015). Araştırma becerileri testinin (ABT) geliştirilmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 12 (2), 305-332.

- Ateş, S. (2002). *Sınıf öğretmenliği ve fen bilgisi öğretmenliği 3. Sınıf öğrencilerinin bilimsel düşünme yeteneklerinin karşılaştırılması*. V. Ulusal Fen Bilimleri ve Matematik Eğitimi Kongresi. 16-18 Eylül 2002. ODTÜ, Ankara.
- Atılğan, H., Kan, A. ve Doğan, N. (2016). *Eğitimde ölçme ve değerlendirme (9. Baskı)*. Ankara: Anı Yayıncılık.
- Aybek, B. ve Çelik, M. (2007). Watson-Glaser Eleştirel Akıl Yürütme Gücü Ölçeği'nin üniversite ikinci, üçüncü ve dördüncü sınıf İngilizce bölümü öğretmen adayları üzerindeki güvenilirlik çalışması. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 16 (1), 101-112.
- Ayvacı, H. Ş. ve Durmuş, A. (2016). Bir başarı testi geliştirme çalışması: Isı ve sıcaklık başarı testi geçerlik ve güvenilirlik araştırması. *Ondokuz Mayıs Üniversitesi Eğitim Fakültesi Dergisi*, 35 (1), 87-103.
- Bao, L., Cai, T., Koenig, K., Fang, K., Han, J., Wang, J., ... Wu, N. (2009). Learning and scientific reasoning. *Science*, 323(5914), 586-587.
- Baran, H. (2018). Açıköğretim sınavlarının faktör yapısının incelenmesi: Uluslararası ilişkiler kuramı-I dersi örneği. *Açıköğretim Uygulamaları ve Araştırmaları Dergisi*, 4 (1), 23-46.
- Benford, R., & Lawson, A. E. (2001). *Relationships between effective inquiry use and the development of scientific reasoning skills in college biology labs*. Master Thesis, Arizona State University.
- Bingöl, A. ve Halisdemir, N. (2017). Üniversite öğrencilerinin temel bilgi teknolojileri dersine yönelik akademik başarı testi geliştirme çalışması. *International Journal of Social Science*, 54, 541-554.
- Brooks, G. P. & Johanson, G. A. (2003). Test Analysis Program. *Applied Psychological Measurement*, 27, 305-306.
- Büyüköztürk, Ş. (2012). *Sosyal Bilimler için Veri Analizi El Kitabı: İstatistik, Araştırma Deseni, SPSS Uygulamaları ve Yorum (17 Baskı)*. Ankara: Pegem Akademi.
- Chen, C. T. & She, H. C. (2015). The effectiveness of scientific inquiry with/without integration of scientific reasoning. *International Journal of Science and Mathematics Education*, 13(1), 1-20.
- Coletta, V. P. & Phillips, J. A. (2005). Interpreting FCI scores: Normalized gain, preinstruction scores, and scientific reasoning ability. *American Journal of Physics*, 73(12), 1172-1182.
- Crocker, S. (2012). *The development of cognition*. Andover, UK: Cengage.
- Demir, E. (2013). Kayıp verilerin varlığında çoktan seçmeli testlerde madde ve test parametrelerinin kestirilmesi: SBS örneği. *Eğitim Bilimleri Araştırmaları Dergisi*, 3 (2), 47-68.
- Demir, N., Kızılay, E. ve Bektaş, O. (2016). 7. Sınıf çözeltiler konusunda başarı testi geliştirme: geçerlik ve güvenilirlik çalışması. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi (EFMED)*, 10 (1), 209-237.
- Ennis, R. H. ve Paulus, D. (1965). *Deductive reasoning in adolescence: Critical thinking readiness in grades 1-12*. Cornell University Cooperative Research Project Report.
- Ennis, R.H., Millman, J., ve Tomko, T.N. (1985). *Cornell Critical Thinking Tests Level X and Level Z— Manual*. Pacific Grove, CA: Midwest Publications.
- Facione, P.A. (1990). *The California Critical Thinking Skills Test—College Level Technical Report 1: Experimental Validation and Content Validity*. Millbrae, CA: California Academic Press.
- Fidan, E. (2013). *İlkokul öğrencileri için matematik dersi sayılar öğrenme alanında başarı testi geliştirilmesi*. Yüksek lisans tezi, Ankara Üniversitesi.
- Fraenkel, J. R. ve Wallen, N.E. (2003). *How to design and evaluate research in education (5th Edition)*. Newyork: McGraw- Hill Companies.
- Geban, Ö., Aşkar, P. ve Özkan, İ. (1992). Effects of computer simulations and problem-solving approaches on high school students. *Journal of Educational Research*, 86 (1), 5-10.
- Gönen, S., Kocakaya, S. ve Kocakaya, F. (2011). Dinamik konusunda geçerliliği ve güvenilirliği sağlanmış bir başarı testi geliştirme çalışması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 8 (1), 40-57.
- Güven, S., & Topbaş, S. (2014). Erken dil gelişimi testi-üçüncü versiyonu'nun (test of early language development-third edition) türkçe'ye uyarlama, geçerlik ve güvenilirlik ön çalışması. *International Journal of Early Childhood Special Education*, 6 (2), 151-176.
- Han, J. (2013). *Scientific reasoning: Research, development and assessment*. Doctoral Thesis, Ohio State University.
- Hanson, S. (2016). The assessment of scientific reasoning skills of high school science students: A standardized assessment instrument. *Illinois State University Theses and Dissertations*, Paper 506.
- Hecht, L.W. (1980). *Validation of the New Jersey College Basic Skills Placement Test*. Technical report. (ERIC Document Reproduction Service No. ED214945.) Princeton, NJ: Educational Testing Service.
- Hogan, K., & Fisherkeller, J. (2005). Dialogue as data: Assessing students' scientific reasoning with interactive protocols. In J. J. Mintzes, J. H. Wandersee, & J. D. Novak (Eds.), *Assessing science understanding: A human constructivist view* (pp. 95-127). Burlington, MA: Elsevier Academic Press.

- Hogan, K., & Maglienti, M. (2001). Comparing the epistemological underpinnings of students' and scientists' reasoning about conclusions. *Journal of Research in Science Teaching*, 38, 663-687.
- iSTAR Assessment. (2010). *iSTAR Assessment: inquiry for scientific thinking and reasoning*. <http://www.istarassessment.org/>
- Karslı, F. ve Ayas, A. (2013). Fen ve teknoloji dersi öğretmen adaylarının bilimsel süreç becerilerinin ölçülmesine ilişkin bir test geliştirme çalışması. *Türk Fen Eğitimi Dergisi*, 10 (2), 66-84.
- Kayılı, G. ve Arı, R. (2015). Wally sosyal problem çözme testi: Geçerlik ve güvenilirlik çalışması. *Uluslararası Eğitim Bilimleri Dergisi*, 2 (3), 51-60.
- Klaczynski, P. A. (2000). Motivated scientific reasoning biases, epistemological beliefs, and theory polarization: A two-process approach to adolescent cognition. *Child Development*, 71, 1347-1366.
- Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science*, 12, 1-48.
- Koslowski, B. (2012). Scientific reasoning: Explanation, confirmation bias, and scientific practice. In G. J. Feist, & M. E. Gorman (Eds.), *Handbook of the psychology of science* (pp. 151-192). New York, NJ: Springer.
- Kuhn, D. (2004). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Blackwell handbook of childhood cognitive development* (pp. 371-393). Maiden, MA: Blackwell.
- Kuhn, D. & Franklin, S. (2006). The second decade: What develops (and how). In W. Damon, R. M. Lerner, D. Kuhn & R. S. Siegler (Eds.), *Handbook of child psychology* (Cognition, perception and language 6th ed., Vol. 2, pp. 953-993). Hoboken, NJ: Wiley.
- Kutlu, Ö. ve Karakaya, İ. (2007). Ortaöğretim kurumları öğrenci seçme ve yerleştirme sınavında kullanılan testlerin faktör yapılarının belirlenmesine ilişkin bir araştırma. *İlköğretim Online*, 6 (3), 397-410.
- Lawson, Anton E. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15, 11-24.
- Lawson, A. E. (2005). What is the role of induction and deduction in reasoning and scientific inquiry? *Journal of Research in Science Teaching*, 42, 716-740.
- Lorenzo-Seva, U. & Ferrando, P.J. (2006). FACTOR: A computer program to fit the exploratory factor analysis model. *Behaviour Research Methods*, 38 (1), 88-91.
- Lorenzo-Seva, U.; Timmerman, M.E.; Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research*, 46(2), 340-364.
- Lorenzo-Seva, U., & Ferrando, P.J. (2013). FACTOR 9.2 A Comprehensive Program for Fitting Exploratory and Semiconfirmatory Factor Analysis and IRT Models. *Applied Psychological Measurement*, 37(6), 497-498.
- Nashwan, Y. (2005). *At-tafkir Al-ilmi wa At-tarbiya Al-amaliyya [Scientific reasoning and practical process]*. Amman, Jordan: Dar Al-Forqan.
- National Research Council (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. H. Quinn, H. A. Schweingruber ve T. Keller (Eds.). Washington, DC: National Academies Press.
- Nawawra, S. (2012). *Scientific reasoning of upper primary school teachers in Bethlehem Governorate* Unpublished Master Thesis, Al-Quds University.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, 44(4), 443-460.
- Osborne, J. (2013). The 21st century challenge for science education: Assessing scientific reasoning. *Thinking Skills and Creativity*, 10, 265-279.
- Özdamar, K. (2016). *Ölçek ve test geliştirme yapısal eşitlik modellemesi*. Eskişehir: Nisan Kitabevi.
- Özmen, B. (2016). *Ortaokul öğrencilerine yönelik bilgi işlemsel düşünme becerileri testinin geliştirilmesi: Geçerlik ve güvenilirlik çalışması*. 4th International Instructional Technologies & Teacher Education Symposium, 6-8 Ekim, Elazığ.
- Sacon, G. (2003). *Incremental value of minimum rank factor analysis*. Groningen:s.n.
- Sadler, T. D., & Zeidler, D. L. (2004). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *Science Education*, 89, 71-93.
- Sandal, M. (2015). *Sıralayıcı ölçme düzeyi için faktör analizi ve bir uygulama*. Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi.
- Schen, M. S. (2007). *Scientific reasoning skills development in the introductory biology courses for undergraduates*. Doctoral Thesis, Ohio State University.
- Schwartz, R. S., Lederman, N. G., & Crawford, B. A. (2004). Developing views of nature of science in an authentic context: an explicit approach to bridging the gap between nature of science and scientific inquiry. *Science Education*, 88, 610-645.

- Sundre, D. L. (2008). *Madison Assessment: The Scientific Reasoning Test, Version 9 (SR-9)*. The Center for Assessment & Research Studies: Harrisonburg.
- Şekercioğlu, G., Bayat, N. Ve Bakır, S. (2014). Fen maddelerini anlama testinin psikometrik niteliklerinin belirlenmesi. *Eğitim ve Bilim*, 39 (176), 447-455.
- Şen, H. C. ve Eryılmaz, A. (2011). Bir başarı testi geliştirme çalışması: Basit elektrik devreleri başarı testi geçerlik ve güvenilirlik araştırması. *Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 8 (1), 1-39.
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston: Pearson/Allyn & Bacon.
- Tekin, H. (2003). *Eğitimde ölçme ve değerlendirme (15. Baskı)*. Ankara: Yargı Yayınevi.
- Tobin, K. G., & Capie, W. (1981). Development and validation of a group test of logical thinking. *Educational and Psychological Measurement*, 41, 413-23.
- Tosun, C. ve Taşkesenligil, Y. (2011). Revize edilmiş Bloom'un taksonomisine göre çözümler ve fiziksel özellikleri konusunda başarı testinin geliştirilmesi: Geçerlik ve güvenilirlik çalışması. *Kastamonu Eğitim Dergisi*, 19(2), 499-522.
- Turgut, M. F. ve Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme (2. Baskı)*. Ankara: Pegem Akademi.
- URL-1.<https://www.act.org/content/act/en/products-and-services/the-act/test-preparation/science-practice-test-questions.html?page=0&chapter=0>. Erişim tarihi: 06.03.2018, 22:19.
- Ülper, H., Çetinkaya, G. ve Bayat, N. (2017). Okuduğunu anlama testinin geliştirilmesi. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 18 (1), 175-187.
- Vass, E., Schiller, D., & Nappi, A. J. (2000). The effects of instructional intervention on improving proportional, probabilistic, and correlational reasoning skills among undergraduate education majors. *Journal of Research in Science Teaching*, 37, 981-995.
- Watson, G & Glaser, M. E. (1964). *Watson-Glaser Critical Thinking Appraisal Manual*, New York: Harcourt, Brace & World Inc.
- Wilkening, F. & Sodian, B. (2005). Scientific reasoning in young children: Introduction. *Swiss Journal of Psychology*, 64(3), 137-139.
- Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, 27, 172-223.