

# Öğretmen Adaylarının Eğitimde Ölçme ve Değerlendirme Dersindeki Kavram Yanılgılarının İncelenmesi

## Examination of Pre-Service Teachers Misconceptions in Measurement and Evaluation Concepts

Sibel AYDOĞAN \*\*

Selahattin GELBAL \*\*\*

### Öz

Bu çalışmanın amacı öğretmen adaylarının eğitimde ölçme ve değerlendirme dersindeki ortak kavram yanılgılarını belirlemektir. Bu amaç doğrultusunda 17 sorudan oluşan iki aşamalı, çoktan seçmeli bir tanı testi geliştirilmiştir. Test puanlarının geçerliği uzman görüşü alınarak, güvenilirliği ise test-tekrar test yöntemi kullanılarak incelenmiştir. Araştırmanın çalışma grubunu 2014-2015 öğretim yılında Hacettepe, Kırıkkale ve Akdeniz Üniversiteleri Eğitim Fakültelerinde öğrenim gören 328 öğretmen adayı oluşturmaktadır. Verilerin analizinde frekans ve yüzde kullanılmıştır. Bulguları yorumlama aşamasında her bir soru için seçilebilecek (4\*4) 16 farklı seçenek ikilisi arasından grubun %10 ve daha fazlası tarafından işaretlenen yanlış kombinasyonlar yaygın yanılğı için bir ölçüt olarak değerlendirmeye alınmıştır. Testteki 17 sorunun analizinden toplam 30 yaygın yanılğı tespit edilmiştir. Çalışmada katılımcıların en çok “güçlük indeksi” ve “normal dağılım” kavramlarında ortak kavram yanılgısına sahip oldukları görülmüştür.

*Anahtar Kelimeler:* eğitimde ölçme ve değerlendirme dersi, iki aşamalı test, kavram yanılgıları, yaygın kavram yanılgısı.

### Abstract

The purpose of this study was to determine pre-service teachers' common misconceptions in measurement and evaluation concepts. In accordance with this objective a two-tier multiple choice test which consists of 17 questions has been developed. The validity of the test was determined by judgements of assessment experts and reliability was examined by test-retest method. The present study was conducted in 2014-2015 academic year with total number 328 pre-service teachers of Hacettepe, Kırıkkale and Akdeniz Universities, Faculty of Education. The obtained data have been analyzed with frequency and percentage. To interpret the findings, the criteria for common misconceptions for each questions was accepted as the 10 percent or more of the options which were chosen among 16 binary alternatives by the group. Thirty common misconceptions were determined among the 17 questions. The most common misconceptions were about the “difficulty index” and “normal distribution” conceptions.

*Keywords:* measurement and evaluation concepts, two-tier test, common misconception.

### GİRİŞ

Başarılı öğrenciler yetiştirmek eğitim sisteminin birincil amaçlarından biridir. Bu amaçta büyük sorumluluğa sahip öğretmenlerin, ölçme ve değerlendirme etkinliklerini sürekli olarak kullanmaları gerekmektedir (Şahin ve Ersoy, 2009). Eğitimin bütün alanlarını ilgilendiren ölçme ve değerlendirme işlemlerinde önemli olan etkinlikleri doğru olarak uygulayabilmektir. Bu sürecin doğru işlemesi için ise ilk şart alanın kapsadığı kavramların doğru yapılandırılmasıdır. Kavramlar, insan zihninde anlaşılan farklı obje ve olguların değişebilen ortak özelliklerini temsil eden bilgi

\* Bu araştırma “Öğretmen Adaylarının Eğitimde Ölçme ve Değerlendirme Dersindeki Kavram Yanılgılarının İncelenmesi” adlı yüksek lisans tezinden üretilmiştir.

\*\* Arş. Gör., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: [sibeldemirbilek07@gmail.com](mailto:sibeldemirbilek07@gmail.com) ORCID ID: [orcid.org/0000-0002-0699-6510](https://orcid.org/0000-0002-0699-6510)

\*\*\* Prof. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, e-posta: [sgelbal@gmail.com](mailto:sgelbal@gmail.com) ORCID ID: [orcid.org/0000-0001-5181-7262](https://orcid.org/0000-0001-5181-7262)

formları olarak adlandırılabilir (Ülgen 2004). İnsanlar kendi düşünce sistemlerince ortaya koydukları kavramlar aracılığıyla düşünebilirler. Ayrıca kavramlar kullanılarak geniş kapsamlı bilgiler daha küçük ve kullanıma elverişli birimler haline getirilebilirler (Senemoğlu, 2003). Kavramların özellikleri sürekli olarak incelenmekte ve yeniden tanımlanmaktadır. Bu sürekli inceleme kavramların tecrübeye dayalı olarak algılanan özellikleri kadar tanımlanabilmesinden kaynaklanmaktadır (Ülgen, 2001). İnsanlar doğumundan ölümüne kadar yeni kavramlar öğrenir ya da var olan kavramlarını yenileri ile değiştirir veya geliştirirler.

Öğrenmelerin başarıya ulaşması için içinde kavramları barındıran yeni bilgiler edinilmesi gerekmektedir. Bu yeni bilgilerin ise var olan bilgiler ile ilişkilendirilmesi ve uygun düzenlemelerin yapılması gerekmektedir. Yeni öğrenilen bir kavram, öğrencilerin zihninde var olan bilgiler ile uygun bir şekilde ilişkilendirilmek yerine, yanlış bir şekilde yapılandırılabilir. Alan yazında kişilerin kavramları bilimsel çevrelerce kabul edilmiş tanımlarından farklı olarak anlamlandırmasına; “kavram yanılgısı”, “ön kavrama”, “alternatif çatı”, “alternatif kavramlar”, “çocukların bilimi” gibi çok çeşitli isimler verilebilmektedir (Treagust, 1988). Bu çalışmada yanlış yapılandırılmış kavramlar “kavram yanılgısı” olarak isimlendirilecektir.

Kavram yanılgısı bireylerin dikkatsizlikleri nedeniyle yaptıkları bir hata veya bilmediklerinden dolayı bir soruya yanlış cevap vermeleri değildir. Kavram yanılgısına sahip olma durumu kişilerin zihinlerinde bir kavramın bilimsel tanımından farklı olarak yapılanması ve değişime dirençli bir şekilde gerekçeleri ile birlikte yerleşmesidir. Kişiler hatalarının doğru olduğunu sebebendirerek açıklıyor ve kendilerinden emin olduklarını söylüyorlarsa o zaman kavram yanılgılarına sahip oldukları söylenebilir. Sahip olunan bütün kavram yanılgıları birer hatadır ama bütün hatalar birer kavram yanılgısı değildir (Eryılmaz ve Sürmeli, 2002). Bu nedenle kavram yanılgılarını diğer hatalardan ayırma süreci zorlaşmaktadır. Bu zorlu süreç kavram yanılgılarını doğru tespit edebilmek amaçlı birçok araç ve yöntem geliştirilmesine neden olmuştur. Doğru tespit amacıyla geliştirilen araç ve yöntemler onları diğer hatlardan ayırt edebilecek nitelikte olmalıdır.

Kavram yanılgılarını belirlemek amacıyla en çok kullanılan yöntem mülakattır. Fakat bu yöntem gerek uygulama sırasında, gerekse bulguları yorumlama aşamasında araştırmacı tarafından karşılaşılabilecek birtakım zorluklar barındırmaktadır (White ve Gunstone, 1992). Bu zorluklar arasında; analizlere araştırmacı sübjektifliğinin karışması ve bu yöntemi kullanmak isteyen araştırmacıların yöntemle ilgili yeterli bilgilerinin olmaması yer almaktadır (Karataş, Köse ve Coştu, 2003). Bu nedenlerle uygulanması ve yorumlanması mülakatlara göre daha kolay olan çoktan seçmeli testler, alan yazında mülakatlardan sonra en çok kullanılan yöntem olmuştur. Fakat çoktan seçmeli test yöntemi ölçülen kavramla ilgili muhakeme konusunda yetersiz kalabilir (Karataş, Köse ve Coştu, 2003). Bu nedenle kavram yanılgılarını ölçmede yetersiz veya hatalı sonuçlar verebilirler. Bu gerekçeler göz önüne alınarak 1988 yılında temelleri Treagust tarafından atılan iki aşamalı kavram testleri günümüzde pek çok araştırmacı tarafından kavram yanılgılarını belirlemek amaçlı kullanılabilir. Alanyazın incelendiğinde bu testlerin iki veya üç aşamalı formlarda kullanıldığı görülmektedir. İki aşamalı testlerin ilk aşaması çoktan seçmeli veya doğru yanlış test formatında olabilmektedir. Bu testlerdeki soruların ikinci aşamasında birinci aşamada seçilen seçeneğin nedeninin belirtilmesi istenmektedir. Eğer üçüncü aşama var ise bu da seçilen seçeneklerin doğruluğundan emin olunup olunmadığının sorulduğu bölümü içermektedir (Eryılmaz, 2010). Bu testlerin ikinci aşaması araştırmacının amacına göre açık uçlu, çoktan seçmeli veya bir seçeneği açık uçlu çoktan seçmeli bir yapıda olabilmektedir. İkinci aşama “çünkü” bağlacıyla başlayarak ilk aşamada seçilen seçeneğin gerekçelendirilmesine imkân tanımaktadır. Bu testler öğrencilerin düşünce yapılarını daha açık bir biçimde ortaya koymasından dolayı, bilinen çoktan seçmeli testlere göre öğrencilerin kavramları nasıl anladıklarını ortaya koymada daha başarılı olabilir. Buna ek olarak iki aşamalı testler, kavram yanılgılarını ölçmede en çok tercih edilen mülakat yöntemine göre de araştırmacıya daha kolay yorumlanabilir sonuçlar sunabilme avantajına sahiptir. Bu özellikleri göz önüne alınan iki aşamalı test, bu çalışmada öğrencilerin kavram yanılgılarını ortaya koyabilmek için tercih edilmiştir.

Kavram yanlışları, öğrencilerin hem ilgili kavramları hem de bu kavramlarla ilişkisi bulunan diğer kavramları öğrenmelerini engellemekte veya geciktirmektedir (Özalp ve Kahveci, 2011). Bu nedenle kavram yanlışlarının tespit edilmesi ve giderilmesi için gerekli çalışmaların yapılması son derece önemlidir. Alanyazın incelendiğinde daha çok fen, matematik ve sağlık alanlarına konu olan kavram yanlışları terimi, ölçme ve değerlendirme alanında çalışan uzmanları da yakından ilgilendirmektedir. Eğitimde ölçme ve değerlendirme dersi, içerisinde istatistik ve matematikle ilişkili kavramlarını da barındıran kapsamlı bir derstir. Ölçme ve değerlendirme ders içeriğinde, doğru yapılandırılmaz ise yanlış yaşanabilecek kavramlar olabilir. Örneğin ne için kullanıldığı öğrenilmeden ezberlenmiş bir standart sapma formülü, öğrencilerin bu kavramla ilgili bilimsellikten uzak yorumlar yapmalarına neden olabilir. Hatta öğrenciler bu bilimsellikten uzak yorumları destekleyici nedenler de öne sürebilirler. Bunlara ek olarak eğitimde ölçme ve değerlendirme ders içeriğinde öğrencilerin karıştırabilecekleri kavram ikilileri de yer alabilmektedir. Buna örnek olarak geçerlik ve güvenilirlik kavramları verilebilir. Bu kavramların tam olarak ne olduğu, birbirleri ile ilişkileri vb. genel olarak öğrencilerin zihinlerinde soru işaretleri bırakabilen konular arasındadır. Bahsedilen örnekler ve daha niceleri ölçme ve değerlendirme uzmanları tarafından incelenebilir. Bu incelemeler sonucunda öğrencilerde yaygın bir şekilde var olan kavram yanlışları tespit edilebilir. Ulusal düzeydeki çalışmalar incelendiğinde ölçme ve değerlendirme dersindeki kavram yanlışları ile doğrudan ilişkili çalışmalar oldukça az olduğu görülmektedir. Doğrudan ilişki çalışmalardan ilki Arık'ın (2006)'da yürüttüğü çalışmasıdır. Arık yaptığı çalışmada öğretmenlerin ölçme ve değerlendirme kavramlarını nasıl algıladıklarını belirlemek üzere kendi geliştirdiği 13 soruluk iki aşamalı kavram testini kullanmıştır. Üztemur (2013)'da ölçme ve değerlendirme dersindeki kavram yanlışlarını belirlediği çalışmada Arık'ın 2006'da geliştirdiği kavram testini kullanmıştır. Bu çalışma kapsamında geliştirilen testin ölçme ve değerlendirme dersindeki kavram yanlışlarını belirlemede kullanılabilecek ikinci test olduğu söylenebilir.

### ***Araştırmanın Amacı***

Araştırmanın amacı eğitimde ölçme ve değerlendirme dersini almış öğretmen adaylarının ders kapsamında belirlenen bazı kavramlardaki yaygın yanlışlarının neler olduğunu iki aşamalı bir test ile belirlemektir.

### **YÖNTEM**

Bu araştırma öğretmen adaylarının kavram yanlışlarını ortaya koymaya yönelik betimsel bir araştırmadır.

### ***Çalışma Grubu***

Çalışma grubu 2014-2015 öğretim yılında Hacettepe, Kırıkkale ve Akdeniz Üniversitelerinin Eğitim Fakültelerinde öğrenim gören öğretmen adayları arasından seçilen 328 öğrenciden oluşmaktadır. Araştırmaya katılan bütün öğrenciler eğitimde ölçme ve değerlendirme dersini çalışmaya katılmadan önce almış ve geçerli not alarak derslerinde başarılı olmuşlardır. Uygulamaya katılan öğrencilerin tümü çalışmaya gönüllü olarak dâhil olmuştur.

### ***Veri Toplama Araçları***

Bu araştırmada veri toplama aracı olarak araştırmacılar tarafından geliştirilen iki aşamalı kavram testi kullanılmıştır. Tanı testinde araştırmacılar tarafından geliştirilen soruların yanında alan yazında daha önce kullanılmış dört soruya da yer verilmiştir. Alan yazındaki bu sorular ilgili araştırmacıdan gerekli izinler alınarak teste dâhil edilmiştir. Bu sorulara, teste dâhil edilme aşamasında birtakım değişiklikler yapılmıştır. Sevimli (2010)'nin uyarılama çalışması yaparak tezinde kullandığı test

istatistik dersindeki kavram yanılgılarını ölçmek amaçlı geliştirilmiş çoktan seçmeli bir testtir. Bu testten alınan sorular iki aşamalı test formuna dönüştürülmüştür.

Tanı testinin geliştirilme aşamasında ilk olarak, test kapsamında yer alacak kavramlar uzman görüşleri ve literatür taraması ile belirlenmiş, ve iki aşamalı testin ilk aşaması oluşturulmuştur. İlk aşaması çoktan seçmeli ikinci aşaması açık uçlu test 78 öğrenciye uygulanarak ikinci aşama için gerekli veriler toplanmıştır. Son aşamada ise tanı testi geliştirilmiştir. Testin puanlarının geçerliliği uzman görüşü alınarak incelenmiştir. Uzmanlar geliştirilen testin ölçme ve değerlendirme dersindeki kavram yanılgılarını ortaya koymak amacıyla kullanılabileceği yönünde görüş bildirmişlerdir. Testin kararlılık anlamında güvenilir sonuçlar üretip üretmediğine ise test-tekrar test yöntemiyle bakılmıştır. Testin hesaplanan güvenilirlik katsayısı 0.81'dir.

Her iki aşaması dört seçenekli kavram testindeki her bir soruda 16 farklı seçenek ikilisi oluşabilmektedir. Bu seçenek ikililerinden biri doğru 15'i yanlıştır. Seçilen her yanlış kombinasyonu kavram yanılgısı olarak değerlendirmek doğru değildir. Kavram yanılgılarının belirlenmesinde Tablo 1'deki sınıflandırılmadan faydalanılmıştır.

Tablo 1. İki Aşamalı Kavram Testlerinde Seçilen Kombinasyonların Sınıflandırılması

		İkinci Aşamının Sonuçları	
		Doğru	Yanlış
Birinci Aşamının Sonuçları	Doğru	Doğru kavram bilgisi	Çoğunlukla yanlış gerekçeli doğru *Nadiren kavram yanılgısı
	Yanlış	Doğru gerekçeli yanlış	*Çoğunlukla kavram yanılgısı Nadiren hata

Kaynak: Eryılmaz, A. (2010). Development and application of three-tier heat and temperature test: Sample of bachelor and graduate students. *Eğitim Araştırmaları – Eurasian Journal of Educational Research*, 40, 17-31.

Tablo 1 incelendiğinde kavram yanılgıların çoğunlukla iki aşamanın da yanlış olduğu seçenek ikilileri arasında olabileceği belirtilmiştir. Bununla birlikte ilk aşamanın doğru ikinci aşamanın yanlış olduğu seçenek ikilileri de nadiren kavram yanılgısı barındırmaktadır. Çalışma kapsamında öğrenciler tarafından işaretlenen seçenek ikilileri Tablo 1 doğrultusunda incelenmiş ve kendi içinde bir tutarlılık barındırılan yanlış ikililer kavram yanılgısı olarak değerlendirmeye alınmıştır. Daha sonra bu ikililerden %10 ve üzerinde olanlar yaygın kavram yanılgısı olarak kabul edilmiş ve bulgu olarak sunulmuştur.

### İşlem

Testin uygulanması için gerekli sınıf ortamı araştırmaya katılan grupların öğretim elemanları tarafından sağlanmıştır. Testin hem iki aşamalı olması hem de 17 sorudan oluşması öğrencilerin cevaplama isteklerini oldukça düşürmüştür. Gerekli motivasyonun sağlanması için araştırmaya katılan öğrencilere testte en yüksek puanı alan kişiye hediye verileceği çalışmanın başında belirtilmiştir. Testin cevaplanması için öğrencilere zaman kısıtlaması verilmemiştir. Bununla birlikte öğrenciler yaklaşık olarak 25-30 dakika arasında testteki soruları tamamlamıştır.

### Verilerin Analizi

On altı seçenek ikilisi oluşan testin verileri 1'den 16'ya numaralandırılarak kayıt edilmiştir. Verilerin analizinde frekans ve yüzde kullanılmıştır.

**BULGULAR**

Bu bölümde çalışmada tespit edilen yaygın yanlışlar (grubun %10 ve daha fazlası tarafından seçilen kombinasyonlar) seçilme yüzdeleri ve soru için geçerli doğru cevap ikilisi ile birlikte verilecektir. Makale içeriğinde testte yer alan bütün sorulara yer verilmeyecektir, bütün sorular ekler bölümünde yer almaktadır. Soruların genel şeklini anlamak amaçlı testin birinci sorusu ve bu sorunun öğrenciler tarafından işaretlenme yüzdeleri örnek olarak sunulmuştur.

Soru 1: Bir çalışmada yemek yeme miktarı ve kilo arasındaki korelasyon +0.90, kandaki tiroit hormonu ve kilo arasındaki korelasyon -0.90 olarak hesaplanıyor.

Bu bulgular doğrultusunda aşağıdaki yorumlardan hangisi yapılabilir?

- A)Tiroit hormonu ve kilo arasında bir ilişki bulunamamıştır.  
 B)Yeme miktarı ile kilo arasındaki ilişki daha güçlüdür.  
 C)Kilo alma ile yemek yeme arasında %90 ilişki vardır.  
 D)Kandaki tiroit hormonu ile kilo arasında yüksek düzeyde bir ilişki bulunmuştur.

Çünkü

- 1)İki değişken arasında, istatistiksel bir ilişki varsa, pozitif bir korelasyon vardır.  
 2)Korelasyon yüzde olarak ifade edilebilir.  
 3) Korelasyon en yüksek ilişki değerine -1.00 ve +1.00 de ulaşır.  
 4)Korelasyon katsayısı -1.00'den 1.00'e doğru gittikçe güçlenir.

Tablo 2'de öğrencilerin birinci soruda seçtikleri kombinasyonların seçilme yüzdeleri yer almaktadır.

Tablo 2. Öğrencilerin Birinci Soruya Verdikleri Cevapların Dağılımı

		Birinci Aşama				
		A	B	C	D*	
İkinci Aşama	1	N	5	20	6	5
		%	1.5 %	6.1 %	1.8 %	1.5 %
	2	N	1	0	24	2
		%	0.3 %	0 %	7.3 %	0.6 %
	3*	N	3	24	5	108*
		%	0.9 %	7.3 %	1.5 %	32.9 %
	4	N	3	86	4	18
		%	0.9 %	26.2 %	1.2 %	5.5 %

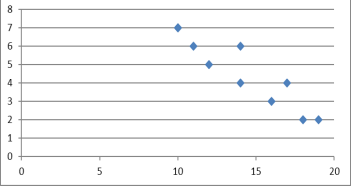
Tablo 2 incelendiğinde doğru cevap ikilisi olan D3 dışında öğrencilerin %10'dan daha fazla işaretledikleri tek kombinasyon B4'tür. Bu çalışmada öğrencilerin %10'dan fazla işaretledikleri seçenek ikilileri yaygın kavram yanlışlığı olabilir yönünde değerlendirmeye alınmış ve toplam 30 seçenek ikilisi yaygın yanlışlık olarak sunulmuştur. Tablo 3'de yaygın yanlışlıklardan önce ilgili sorudaki doğru kavram ikilisi verilmiştir. Daha sonra 17 soruda tespit edilen 30 yaygın yanlışlık seçilme yüzdesine göre büyükten küçüğe sıralanarak sunulmuştur.

Tablo 3. Doğru Kavram İkili, Yaygın Yanılgılar ve Seçilme Yüzdeleri

Soru Numarası	Seçilme yüzdesi	Doğru kavram ikilisi	Seçilme yüzdesi	Yaygın yanılgılar															
17	%14,6	<table border="1"> <thead> <tr> <th>Soru</th> <th>Güçlük</th> <th>Ayrıtedicilik</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0,75</td> <td>0,80</td> </tr> <tr> <td>2</td> <td>0,25</td> <td>0,20</td> </tr> <tr> <td>3</td> <td>0,75</td> <td>0,20</td> </tr> <tr> <td>4</td> <td>0,25</td> <td>0,80</td> </tr> </tbody> </table> <p>“Tablodaki istatistiklere göre en ayırt edici ve zor olan soru 4’üncü sorudur. Çünkü sorunun hem ayırt edici hem de zor olması için ayırt edicilik gücü 1’e, güçlük indeksi 0’a yaklaşmalıdır.”</p>	Soru	Güçlük	Ayrıtedicilik	1	0,75	0,80	2	0,25	0,20	3	0,75	0,20	4	0,25	0,80	%50,9	“Tablodaki istatistiklere göre en ayırt edici ve zor olan soru 1’inci sorudur. Çünkü sorunun hem ayırt edici hem de zor olması için ayırt edicilik gücü 1’e, güçlük indeksi 1’e yaklaşmalıdır.”
Soru	Güçlük	Ayrıtedicilik																	
1	0,75	0,80																	
2	0,25	0,20																	
3	0,75	0,20																	
4	0,25	0,80																	
6	%14,3	“İlköğretim 3. sınıf öğrencilerinin zekâ düzeylerinin normal dağılım göstermesini beklerim; çünkü bir veri grubunun normal dağılım göstermesi frekans eğrisinin çan şeklinde ve simetrik olmasına bağlıdır.”	%50,6 %9,8	“İlköğretim beşinci sınıf öğrencilerinin yaşlarının normal dağılım göstermesini beklerim; çünkü bir veri grubunun normal dağılım göstermesi frekans değerlerinin bir değer etrafında toplanmasına bağlıdır.” “İlköğretim beşinci sınıf öğrencilerinin yaşlarının normal dağılım göstermesini beklerim; çünkü bir veri grubunun normal dağılım göstermesi her bir alt grubun frekans değerlerinin eşit olmasına bağlıdır”															
15	%42,7	“-0,85 düzeyindeki korelasyon yüksek bir ilişkiyi gösterir. Çünkü ilişkinin gücü yorumlanırken korelasyon katsayısının işareti göz önünde bulundurulmaz.”	%37,2	“-0,85 olarak hesaplanan korelasyon katsayısı düşük bir ilişkiyi gösterir. Çünkü korelasyon eksi çıktığı için aralarında zayıf bir ilişki olduğu sonucuna ulaşılabilir.”															
10	%30,8	“Ayşe öğretmenin her öğrenciye almış olduğu puanın %10 fazlasını vermesi ölçme sonuçlarına sistematik hata karıştırmaktır. Çünkü ölçme sonuçları sabit oranlarda değişmiştir.”	%36,1	“Her öğrenciye almış olduğu puanın %10 fazlasını vermek sabit hatadır çünkü bütün ölçme sonuçları aynı miktarda etkilenmiştir.”															
1	%32,9	“Kandaki tiroit hormonu ile kilo arasında bulunan -0.90 korelasyon yüksek düzeyde bir ilişkiyi gösterir. Çünkü korelasyon en yüksek ilişki değerine -1.00 ve +1.00’de ulaşır.”	%26,2	“+0.90 korelasyon, -0.90’a göre daha güçlü bir ilişkiyi gösterir. Çünkü korelasyon katsayısı -1.00’den +1.00’e doğru gittikçe güçlenir.”															
14	%4,6	“Maddenin güçlüğü 1’e çok çok yakın ise madde ayırt ediciliği 0’a çok yakın olur. Çünkü maddeyi doğru cevaplayanların çok olması, iyi ve iyi olmayan öğrencileri ayırt etmeyi zorlaştırır.”	%24,4	“Madenin güçlüğü 1’e çok çok yakın ise ayırt ediciliği 1’e çok yakın olur. Çünkü maddeyi doğru cevaplayanların çok az olması nedeni ile zayıf öğrenciler daha iyi ayırt edilir.”															
12	%29,9	“Tutarlılık, duyarlılık ve kararlılık kavramları ölçme aracının güvenilirlik özelliği ile ilişkilidir. Çünkü bu kavramlar teste daha az hata karıştığının göstergesidir.”	%23,5 %11,6	“Tutarlılık, duyarlılık ve kararlılık kavramları ölçme aracının geçerlik özelliği ile ilişkilidir. Çünkü bu kavramlar testin ölçmek istediği davranışları ölçtüğünün göstergesidir.” “Tutarlılık, duyarlılık ve kararlılık kavramları ölçme aracının güvenilirlik özelliği ile ilişkilidir. Çünkü bu kavramlar testin ölçmek istediği davranışları ölçtüğünün göstergesidir.”															



Tablo 3. Doğru Kavram İkili, Yaygın Yanılgılar ve Seçilme Yüzdeleri (Devam)

Soru Numarası	Seçilme yüzdesi	Doğru kavram ikilisi	Seçilme yüzdesi	Yaygın yanılgılar																		
13	%6,4	A ve B grubuna uygulanmış iki ayrı testin istatistikleri <table border="1"> <thead> <tr> <th></th> <th>Puan Aralığı</th> <th>Standart Sapması</th> <th>Ortalaması</th> </tr> </thead> <tbody> <tr> <td>A</td> <td>0-100</td> <td>10</td> <td>50</td> </tr> <tr> <td>B</td> <td>0-10</td> <td>2</td> <td>5</td> </tr> </tbody> </table> <p>“Tabloya göre B testini alan grup A testini alan gruba göre ölçülen özellik bakımından daha çok farklılık göstermiştir. Çünkü iki grubun farklılaşması karşılaştırılırken standart sapma ve ortalama birlikte dikkate alınır.”</p>		Puan Aralığı	Standart Sapması	Ortalaması	A	0-100	10	50	B	0-10	2	5	%22,6  %13,1	“A testini alan grup B testini alanlara göre daha çok fark göstermiştir. Çünkü iki grubun farklılaşması karşılaştırılırken standart sapma dikkate alınır.” “Verilen istatistikler yeterli değildir. Çünkü iki grubun farklılaşması karşılaştırılırken basıklık katsayısı verilmeden yorum yapılamaz.”						
	Puan Aralığı	Standart Sapması	Ortalaması																			
A	0-100	10	50																			
B	0-10	2	5																			
4	%17,4	“3, 4, 5, 6, 6, 8, 10, 12, 19, 26, 83 verilerinden oluşan kümenin ortanca ile temsil edilmesi en uygunu olur. Çünkü veri kümesinin yanlış değerlendirilmesine neden olacak uç değerlerin etkisini azaltmalıyız.”	%20,1	“3, 4, 5, 6, 6, 8, 10, 12, 19, 26, 83 verilerinden oluşan kümenin aritmetik ortalama ile temsil edilmesi en uygunu olur. Çünkü bütün değerlerin kullanılarak hesaplandığı bir istatistik seçmeliyiz.”																		
16	%12,8	<table border="1"> <thead> <tr> <th></th> <th>A</th> <th>B</th> <th>C*</th> <th>D</th> <th>N</th> </tr> </thead> <tbody> <tr> <td>ÜST GRUP</td> <td>11</td> <td>30</td> <td>15</td> <td>4</td> <td>50</td> </tr> <tr> <td>ALT GRUP</td> <td>15</td> <td>11</td> <td>4</td> <td>20</td> <td>50</td> </tr> </tbody> </table> <p>“Bu soruda en iyi çalışan çeldirici D seçeneğidir. Çünkü alt grup-üst grup farkının fazla olduğu seçenek en iyi çeldiricidir.”</p>		A	B	C*	D	N	ÜST GRUP	11	30	15	4	50	ALT GRUP	15	11	4	20	50	%19,8  %17,1	“Bu soru için en iyi çalışan çeldirici B seçeneğidir. Çünkü üst gruptan en çok kişiyi çeken şık en iyi çeldiricidir.” “Bu soru için en iyi çalışan çeldirici B seçeneğidir. Çünkü toplamda en çok işaretlenen şık en iyi çeldiricidir.”
	A	B	C*	D	N																	
ÜST GRUP	11	30	15	4	50																	
ALT GRUP	15	11	4	20	50																	
8	%22,6	“Metrenin birimleri yapaydır. Çünkü ölçme aracı birimleri ortak kararlar belirlenmiştir.”	%14,3  %17,4	“Metre aracının birimi doğaldır. Çünkü aracın sıfır noktası mutlak sıfırdır.” “Metre aracının birimi doğaldır. Çünkü ölçme aracı ölçmek istediği özelliği direk olarak ölçer.”																		
2	%25,6	“Düşük puandan yüksek puana doğru sıralamada grubun 95’inci yüzdeliğinde bulunan öğrenci grubun %95’ini geride bırakmıştır. Çünkü yüzdeler küçükten büyüğe sıralanmış verilerin belli bir yüzdesini altında bırakan noktadaki değerlerdir.”	%17,1	“Düşük puandan yüksek puana doğru sıralamada grubun 95’inci yüzdeliğinde bulunan öğrenci grubun %5’ini geride bırakmıştır. Çünkü düşük puandan yüksek puana doğru sıralanmış veri grubunda 95’inci yüzdeler ifadesi ölçümlerin %95’ini üstünde %5’ini altında bulunduran noktadaki değerlerdir.”																		
5	%17,4	 <p>%17,4 “Dağılım grafiğinden yararlanarak X ve Y arasındaki korelasyon katsayısını -0,9 olarak tahmin edebiliriz. Çünkü veriler arasında güçlü bir ilişki vardır.”</p>	%16,2  %13,7  %11,3  %10,7	“X ve Y arasındaki korelasyon katsayısını 0.0 olarak tahmin edebiliriz. Çünkü veriler arasında ilişki yoktur.” “X ve Y arasındaki korelasyon katsayısını 0.8 olarak tahmin edebiliriz. Çünkü veriler arasında güçlü bir ilişki vardır.” “X ve Y arasındaki korelasyon katsayısını -0.9 olarak tahmin edebiliriz. Çünkü veriler arasında zayıf bir ilişki vardır.” “X ve Y arasındaki korelasyon katsayısını -0.1 olarak tahmin edebiliriz. Çünkü veriler arasında zayıf bir ilişki vardır.”																		

Tablo 3. Doğru Kavram İkilisi, Yaygın Yanılgılar ve Seçilme Yüzdeleri (Devam)

Soru Numarası	Seçilme yüzdesi	Doğru kavram ikilisi	Seçilme yüzdesi	Yaygın yanılgılar
9	%11,9	%11,9 “Uygulama sonunda sınav kâğıtlarına 0 ile 100 arasında puanlar verme işlemi ölçmedir. Çünkü gözlem sonuçları sayılarla ifade edilmiştir.”	%14,2 %14,2	“Uygulama sonunda sınav kâğıtlarına 0 ile 100 arasında puanlar verme işlemi değerlendirmedir. Çünkü bu işlemde ölçümler ölçütlerle karşılaştırılarak karara varılmıştır.” “Uygulama sonunda sınav kâğıtlarına 0 ile 100 arasında puanlar verme işlemi ölçüt belirlemedir. Çünkü ölçme kuralı belirlenmiştir.”
3	%51,2	“Serideki en yüksek puanın değişmesi ile medyan değişmez, standart sapma değişir. Çünkü bir serideki en yüksek puanın değişimi medyanyı etkilemez.”	%13,1 %11,6	“Serideki en yüksek puanın değişmesi ile bir tek ranj değişmez. Çünkü bir serideki en yüksek puanın değişiminden bir tek ranj etkilenmez.” “Serideki en yüksek puanın değişmesi ile medyan değişir, standart sapma değişmez. Çünkü bir serideki en yüksek puanın değişimi standart sapmayı etkilemez.”
11	%21,6	“Geçerli her test güvenilirdir. Çünkü testin ölçmek istediği davranışı ölçebilmesi için önce hatalardan arınması gerekir.”	%13,1 %11,9	“Geçerlik ve güvenilirlik birbirini etkileyen kavramlar değildir. Çünkü bunlar bağımsız kavramlardır.” “Güvenilir her test geçerlidir. Çünkü hatalardan arınık bir test ölçmek istediği özelliği ölçebilir.”
7	%31,4	“50 ölçüm ile hesaplanan standart sapma değeri -0,8 çıkmış ise standart sapma yanlış hesaplanmıştır; çünkü standart sapma negatif değer alamaz.”	%12,5 %12,5 %10,7	“Verilerin standart sapması -0,8 ise ölçümlerin çoğu negatiftir. Çünkü standart sapma ölçümler negatif iken negatif değer alır.” “Verilerin standart sapması -0,8 ise ölçümlerin tümü ortalamadan daha küçüktür. Çünkü standart sapma ortalamadan farkların toplamıdır.” “Verilerin standart sapması -0,8 ise ölçümlerin tümü ortalamadan daha küçüktür. Çünkü standart sapma -1 ile +1 arasında değerler alabilir.”

Tablo 3 incelendiğinde 17 soruda 30 yaygın yanılığın tespit edildiği görülebilir. Test kapsamında ölçülen kavramlardan en fazla yanılığın yaşananlar maddenin güçlüğü, normal dağılım, korelasyon ve hata türleri olmuştur.

## SONUÇLAR ve TARTIŞMA

Bu çalışmada öğretmen adaylarının ölçme değerlendirme dersi ile ilgili kavram yanılgılarının belirlenmesi amaçlanmıştır. Bu bağlamda toplam 17 sorudan oluşan iki aşamalı kavram yanılgıları testi geliştirilmiştir. Araştırma kapsamında geliştirilen bu testin güvenilirlik ve geçerlik çalışmaları yapılmış ve testin öğretmen adaylarının ölçme değerlendirme ile ilgili kavram yanılgılarını ortaya çıkaran güvenilir ve geçerli bir ölçme aracı olduğu ortaya konmuştur.

Arık (2006) tarafından geliştirilen 13 soruluk kavram testi ilk aşaması iki, ikinci aşaması 4 seçenekli bir formda hazırlanmıştır. Üztemur (2013) da çalışmasında Arık'ın geliştirdiği testi kullanmıştır. Bu çalışmalarda ilk aşamada yanlış, ikinci aşama doğru seçim yapan bireylerin kavram yanılığısına sahip



olduğu belirtilmiştir. Bu doğrultuda iki çalışmada da en çok “düzeltme puanı” kavramında kavram yanlışlığı olduğu belirtilmiştir. Bu çalışmalarda örneğin; “4 yanlışın 1 doğruyu götürmesi işlemi sonucunda elde edilen puana” birinci aşamada “düzeltme puanı” yerine “ham puanı”, ikinci aşamada ise bu işlemin yapılma nedeninin “test puanlarının şans başarısından arındırılması” şeklinde işaretleme yapanlar kavram yanlışlığına sahip olarak şeklinde kategorileştirilmiştir. Alanyazında ölçme ve değerlendirme alanında yer alan çalışmalardan farklı olarak bu çalışmada, bu tür yanlışlar daha çok kavramların karıştırılması olarak ele alınmış, her iki aşamada da kendi içinde tutarlılık barındıran yanlış ikililer kavram yanlışlığı olarak değerlendirilmiştir. Kavram yanlışlığının bu şekilde değerlendirilmesinde Tablo 1’deki sınıflandırmadan faydalanılmıştır.

Hacettepe, Kırıkkale ve Akdeniz Üniversitelerinde öğrenim görmekte olan Eğitim Fakültesi öğrencileri ile yürütülen çalışmanın sonuçları, öğretmen adaylarının bu alandaki birçok kavramda zorluk yaşadıklarını ortaya çıkarmıştır. Testte kavram yanlışlığı içeren ve katılımcıların en az %10’u tarafından işaretlenen toplam 30 seçenek ikilisi tespit edilmiştir.

Test genel olarak incelendiğinde öğrencilerin en çok güçlük indeksi kavramında ve normal dağılım gösteren veri kümesini belirlemede yanlış yaşadıkları söylenebilir. Öğrencilerin güçlük indeksi kavramında yanlış yaşamalarının nedeni, kavramın Türkçe karşılığının yanıltıcı olması olabilir. ‘Güçlük indeksi 0’den 1’e doğru gittikçe sorunun zorlaşacağı’ kulağa birçok kişi tarafından mantıklı gelmektedir. Fakat bu yanlış bir bilgidir. Bu kavramın ismi sorunun ‘kolaylık indeksi’ olsa belki soru birçok öğrenci tarafından doğru cevaplanmış olacaktır. Bu kavramın Türkçe ifadesinin uzmanlar tarafından tekrar gözden geçirilmesi yararlı olabilir. Çalışmaya katılan öğrencilerin %50,6’sı ilköğretim beşinci sınıf öğrencilerinin yaşlarının, bir değer etrafında toplanması gerekçesiyle normal dağılım göstermesini beklemektedir. Bu noktada normal dağılımla ilgili somut örnekler işe koşulabilir. Bu sayede öğrencilerin kavramı daha net anlaması sağlanabilir. Korelasyon kavramını ölçen sorularda ortak yanlışlar genel olarak eksi korelasyonun ilişki belirtmediği veya artı korelasyonun eksiye göre daha güçlü bir ilişkiyi gösterdiği yönündedir. Korelasyon grafiğini okuma becerisinin ölçüldüğü 5’inci soruda ise öğrencilerin grafiği okumada yetersiz kaldıkları gözlenmiştir. Eksi 0,92 düzeyinde korelasyonu temsil edecek şekilde çizilen grafikte öğrencilerin %33,4’ü grafiği zayıf veya ilişki yok şeklinde yorumlamıştır. Grubun %13,7’si ise veriler arasındaki güçlü ilişkiyi fark etmiş, fakat ilişkinin yönünü pozitif olarak değerlendirerek yine grafiği okumada yetersiz kalmışlardır. Bu yanlışların önüne bilgisayar desteğini sınıf ortamında kullanarak ve öğrencilere farklı şekillerde verilen veri gruplarının grafikleri çizdirilerek geçilebilir. Öğrencilerin sabit ve sistematik hata kavramlarını karıştırdıkları da testten elde edilen bir sonuçtur. Öğrenciler merkezi eğilim ve dağılım ölçülerinde de yaygın yanlışlar yaşamışlardır. Öğrencilerde yaygın olarak yer alan bir yanlış da bir maddeyi doğru cevaplayanların çok az olması sonucunda o maddenin çok iyi ayırt edici olacağı yönündedir. Ayrıca testte yer alan ve kapsamı geçerlik ve güvenilirlik kavramların özelliklerini ölçen sorularda öğrencilerin bu kavramlara ait özellikleri karıştırdıkları söylenebilir. Katılımcılar puanlama sistemleri farklı iki sınavı, bağıl değişkenlik katsayısı ile yorumlama konusunda da yeterli başarıyı gösterememişlerdir. Öğrencilerin 5’te 1’i veri kümesinde uç değerler olsa dahi aritmetik ortalamanın veri kümesini en iyi temsil eden merkezi eğilim ölçüsü olduğunu düşünmektedir. On altıncı soruda en iyi çalışan çeldirici özelliklerini yansıtmada yaygın yanlışlar yaşayan öğrenciler, üst grup- alt grup yöntemi ile madde ayırt ediciliğini belirlemede de uygun seçenekleri işaretlemede yetersiz kalmışlardır. Testten elde edilen bir diğer sonuçta öğrencilerin ölçme ve değerlendirme kavramlarını karıştırabilecekleri yönündedir. Ayrıca katılımcılar merkezi dağılım ölçülerinden standart sapma kavramında örneğin standart sapmanın eksi çıkabileceği şeklinde yanlışlar göstermişlerdir.

Çalışmanın bulgularında en az doğru cevaplanan sorular ile en yaygın kavram yanlışlığını barındıran soruların eşleşmediği görülmektedir. Bu noktadan hareketle öğrencilerin en az doğru yaptıkları sorularda en yaygın kavram yanlışlığı yaşadıklarını savunmanın yanlış olduğu çalışma kapsamında ulaşılan bir sonuçtur. Bu tür bir bulguya çoktan seçmeli testlerle ulaşılamaz. Bir öğrencinin tesadüfi hata ya da dikkat dağınıklığından dolayı çoktan seçmeli bir soruyu yanlış cevaplaması olağandır. Fakat iki aşamalı testlerdeki gibi birbiri ile ilişkili iki yanlış cevap, öğrencinin kavram yanlışlığına sahip olduğunu savunmak için daha güçlü deliller sunar. Bunun

yanında iki aşamalı testlerde doğru cevap verilmesi kavramın büyük olasılıkla doğru yapılandırıldığına da kanıtı sayılabilir. Bu noktalar dikkate alındığında iki aşamalı testlerin çoktan seçmeli testlere göre öğrencilerin kavramları nasıl algıladıklarını anlamada daha elverişli olduğu söylenebilir.

Çalışmada kullanılan iki aşamalı kavram testinin her aşaması 4 seçenekli çoktan seçmeli bir formda hazırlanmıştır. Bu şekilde her bir soru için yapısı standart olan bir testin bulgularını karşılaştırmanın anlamlı olacağı söylenebilir. Ayrıca testin her aşamasının 4 seçenekli olması öğrencilere iki aşamada 16 farklı seçenek kombinasyonu sunmuştur. Bu kadar seçenek arasından en az %10 tarafından seçilen yanlış kombinasyonların neredeyse hepsinin kavram yanılgısı özelliği taşıdığı görülmüştür. Chen, Lin ve Lin (2002)'nin lise öğrencilerinin düzlem aynalarla ilgili kavram yanılgılarını ölçmeyi amaçladıkları çalışmalarında da %10 dan fazla işaretlenen seçenek ikilileri olası kavram yanılgısı olarak değerlendirilmeye alınmış ve bu seçenek ikililerinin çoğunlukla kavram yanılgısı olduğu görülmüştür.

Kavram yanılgıların tespit edilmesi ve giderilmesi noktasında öğretmenlere büyük görev düşmektedir. Her konunun sonunda kavram yanılgılarını ortaya koyacak kısa testler uygulanabilir ve bu sayede kavram yanılgıları saptanabilir. Bu saptamalar doğrultusunda düzenlenecek eğitim ortamları ile yanılgıların giderilmesi kolaylaşabilir.

İki aşamalı testlerin hazırlanması ve geliştirilmesi zorlu bir süreçtir. Fakat test geliştirildikten sonra uygulanması ve bulgularının analiz edilmesi oldukça hızlı olabilir. Alan uzmanları öğrencilerin kavramları nasıl algıladıklarını ortaya koyacak geçerlik ve güvenilirliği yüksek iki veya üç aşamalı testler geliştirebilirler.

İki aşamalı testlerin kavram yanılgılarını belirlemenin yanında eğitimin birçok alanında kullanımının da yararlı olacağını düşünülebilir. Çünkü bu testler yüzeysel öğrenmelerden ziyade anlamlı öğrenmelere odaklanmaktadır. Yani bu testlerin doğruları da, yanlışları da araştırmacıya değerli sonuçlar sunmaktadır denebilir.

Araştırmada örneklem uygun örnekleme yöntemi ile oluşturulduğu için bölümler arası homojen bir dağılım sağlanamamıştır, bu konudaki başka araştırmalarda örneklem evreni daha iyi temsil edilecek şekilde seçilebilir. Araştırmaya dâhil edilemeyen eğitim fakültesi bölümleri mevcuttur, o bölümler başka çalışmalarda incelenebilir.

Araştırmada kullanılan test eğitimde ölçme ve değerlendirme dersi kapsamındaki kavramların birçoğunu kapsayacak şekilde geliştirilmiştir. Bu konuyla ilgilenecek araştırmacılar konuları bölerek daha kısa ve katılımcıları daha kolay motive edebilecekleri testler geliştirebilirler. Konuları bölmek soruları birbiri ile daha rahat ilişkilendirip daha kapsamlı sonuçlar elde edilmesine yardımcı olabilir.

Testten özellikle yaygın kavram yanılgılarını ortaya koyan sorular seçilip üzerine yeni sorular eklenebilir. Bu işlem öğretmen adaylarının ölçme ve değerlendirme dersinde yaşadıkları zorlukları daha geniş bir perspektiften görmeye imkân sağlayabilir.

Çalışmada bazı kavram yanılgıları %35'in üzerinde tespit edilmiştir. Bu yanılgıların derinlemesine incelenmesi başka bir araştırmanın konusu olabilir. Yanılgıların sebepleri ve çözüm önerileri incelenebilir.

## KAYNAKÇA

- Arık, S. R. (2006). *İlköğretim öğretmenlerinin ölçme ve değerlendirme alanındaki kavram yanılgılarının belirlenmesi* (Yayımlanmamış Yüksek Lisans Tezi, Ankara). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Eryılmaz, A. (2010). Development and application of three-tier heat and temperature test: Sample of bachelor and graduate students. *Eğitim Araştırmaları – Eurasian Journal of Educational Research*, 40, 17-31.
- Eryılmaz, A. ve Sürmeli, E. (2002, Eylül). *Üç aşamalı sorularla öğrencilerin ısı ve sıcaklık konularındaki kavram yanılgılarının ölçülmesi*. V. Ulusal Fen Bilimleri ve Matematik Kongresi, ODTÜ, Ankara.

- Chen, C. C., Lin, H. S., & Lin, M. L. (2002). Developing a two-tier diagnostic instrument to assess high school students' understanding-the formation of images by a plane mirror. *Proceedings of National Science Council, 12*(3), 106-121.
- Karataş, F. Ö., Köse, S. ve Coştu, B. (2003). Öğrenci yanılgılarını ve anlama düzeylerini belirlemede kullanılan iki aşamalı testler. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi, 13*(1), 54-69.
- Özalp, D. ve Kahveci, A. (2011). Maddenin tanecikli yapısı ile ilgili iki aşamalı tanılayıcı soruların ontoloji temelinde geliştirilmesi. *Milli Eğitim Dergisi, 40*(191), 135-155.
- Senemoğlu, N. (2003). *Gelişim öğrenme ve öğretim*. Ankara: Gazi Kitabevi.
- Sevimli, N. E. (2010). *Matematik öğretmen adaylarının istatistik dersi konularındaki kavram yanılgıları; istatistik dersine yönelik öz yeterlilik inançları ve tutumlarının incelenmesi* (Yüksek Lisans Tezi, Marmara Üniversitesi Eğitim Bilimleri Enstitüsü, İstanbul). <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- Şahin, Ç. ve Ersoy, E. (2009). Sınıf öğretmeni adaylarının yeni ilköğretim programındaki ölçme-değerlendirme konusundaki yeterlilik düzeylerine ilişkin algıları. *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi, 18*(2), 363-386.
- Treagust, D. F. (1988). Development and use of diagnostic tests to evaluate students' mis-conception in science. *International Journal of Science Education, 10*(2), 159-169.
- Ülgen, G. (2001). *Kavram geliştirme: Kurallar ve uygulamalar*. Ankara: Pegem.
- Ülgen, G. (2004). *Kavram geliştirme kuramlar ve uygulamalar* (4. Baskı) Ankara: Nobel.
- Üztemur, S. S. (2013). *Sosyal bilgiler öğretmenlerinin ölçme ve değerlendirme alanındaki kavram yanılgıları ve öz-yeterlilik inançlarının incelenmesi* (Yüksek Lisans Tezi). Celal Bayar Üniversitesi, Manisa. <http://tez2.yok.gov.tr/> adresinden edinilmiştir.
- White, R. T., & Gunstone, R. F. (1992). *Probing understanding*. London: The Falmer Press.

## EXTENDED ABSTRACT

### Introduction

The ability to implement activities properly is an important part of both the assessment as well as evaluation processes within all areas of education. The first condition in order for these processes to function properly is (thus) the proper structuring of concepts within a given field. Newly acquired concepts may however be improperly structured instead of either being associated with the knowledge present in students' minds, or instead of the application of proper configurations. In this study, these improperly structured concepts have been labelled as "misconception(s)"

The situation of having misconceptions entails the structuring of a concept as being from that concept's academic/scientific definition, together with and with the implanting of facts in a manner that is resistant to chance into the minds of individuals. In essence, all misconceptions are mistakes, however not all mistakes are misconceptions (Eryılmaz and Sürmeli, 2002). Consequently, the process of distinguishing misconceptions from other mistakes becomes more challenging. This challenging process has led to the development of many tools and methods for accurately identifying misconceptions. Methods that are developed with the aim of accurate identification should be of a sufficient level of quality in order to distinguish misconceptions from other forms of mistakes.

Two-tier concept tests, as developed in 1988 by Treagust, are now used by many researchers in order to assess misconceptions. The second tier of this test asks participants to specify their reasons for in the choices they made in the first tier. These tests may be more successful than commonly known multiple-choice tests with respect to revealing how students understand concepts, as they reveal the thinking patterns of students in a more explicit manner. Moreover, two-tier tests carry the advantage of offering researchers results that are easy to interpret.

Misconceptions either hinder or delay students' acquisition of relevant concept as well as other concepts that correlate with these concepts (Özalp and Kahveci, 2011). Therefore, studies regarding the determining and eliminating of misconceptions of utmost importance. An assessment and evaluation course within education is a comprehensive course that also encompasses concepts that are part of statistics and mathematics. There may be concepts that should be properly structured correctly within the course content that, if not, could otherwise lead to misconceptions. For example, a standard deviation formula memorized without reason may lead to students interpret concepts in an

unscientific manner. Students may offer reasons supporting the logic behind such unscientific interpretations. In addition, the course in question may contain concept pairs that students can easily confuse. The concepts of validity and reliability are good examples of this. What these concepts are exactly and how they are related to each other etc., are among any number of things that leave students with more questions than answers. These examples alongside many others can be investigated by assessment and evaluation experts. Moreover, misconceptions that are commonly found in and among students may be identified in the results of analysis.

### ***Method***

This is a descriptive study aimed at revealing misconceptions held by student teachers. The study group was formed by selecting 328 prospective student teachers studying in the respectful education faculties of Hacettepe, Kırıkkale and Akdeniz University during the 2014-2015 academic year.

We used the two-tier diagnostic test as a data collection tool developed by researchers in this study. 16 different combinations could be formed for each question in the concept test, which consists of four choices in the first tier as well as four choices at the second tier. One of these choice combinations is correct, while the remaining 15 are wrong. It is not right to consider every wrong choice combination as being a misconception since a choice combination needs to have consistency within itself in order to be identified as a misconception. Therefore, we determined the combinations chosen by the participants as matching misconceptions. We then considered those that were 10% or above as being common misconceptions, and then presented these as part of the findings.

### ***Results and Discussion***

We determined that a total 30 choice combinations contained misconceptions (10% or above) from the 17 questions contained within the test.

Here, one can note that in the overall examination of the test that the students most commonly have misconceptions when it comes to the concepts of difficulty index and determining data set with normal distribution. Common wrong choices in the questions assessing the concept of correlation generally states that negative correlation does not indicate any relationship whatsoever, or positive correlation shows stronger relation when compared to negative correlation. Students were observed as not being fully capable of reading the graphic in the fifth question assessing correlation graphic reading skills. This misconception may be avoided by using computer assistance in classroom, as well as by making students plot graphics of data sets provided in various ways. Another misconception commonly found among students was that an item is more distinctive when answered correctly by few. Moreover, it can be noted that students confuse attributes of the concepts of validity and reliability in the questions assessing knowledge of these concepts. Students furthermore demonstrated misconceptions when it came to standard deviation, which is one of the central distribution measurements—this can be seen as being negative.

It is very normal for a student to choose a wrong answer due to a lack or distraction of attention. However, two wrong answers that correlate with one other provide us with stronger evidence that the student has a misconception. Similarly, answering two-tier tests correctly may be evidence for the proper structuring of a/the concept in question. Given these facts, it can be noted that two-tier tests are better in understanding how students comprehend concepts as compared to multiple choice tests.

Four choices at each tier of the test provided 16 different choice combinations at two tiers to the students. We observed that a minimum of 10% of wrong combinations made by students bears the attributes of misconception.

Instructors have an important role in determining and eliminating misconceptions. Short tests to reveal misconceptions can be applied at the end of each subject. Misconceptions can thus be

identified/determined. The elimination of misconceptions can be facilitated by means of the education environments improved in line with such determinations.

## EKLER

### Ek 1

#### EĞİTİMDE ÖLÇME VE DEĞERLENDİRME KAVRAM TESTİ

1. Bir çalışmada yemek yeme miktarı ve kilo arasındaki korelasyon +0.90, kandaki tiroit hormonu ve kilo arasındaki korelasyon -0.90 olarak hesaplanıyor.

Bu bulgular doğrultusunda aşağıdaki yorumlardan hangisi yapılabilir?

- A) Tiroit hormonu ve kilo arasında bir ilişki bulunamamıştır.
- B) Yeme miktarı ile kilo arasındaki ilişki daha güçlüdür.
- C) Kilo alma ile yemek yeme arasında %90 ilişki vardır.
- D) Kandaki tiroit hormonu ile kilo arasında yüksek düzeyde bir ilişki bulunmuştur.

Çünkü

- 1) İki değişken arasında, istatistiksel bir ilişki varsa, pozitif bir korelasyon vardır.
- 2) Korelasyon yüzde olarak ifade edilebilir.
- 3) Korelasyon en yüksek ilişki değerine -1.00 ve +1.00 de ulaşır.
- 4) Korelasyon katsayısı -1.00'den 1.00'e doğru gittikçe güçlenir.

2. Bir öğrenci matematik sınavı için hesaplanan puanı ile düşük puandan yüksek puana doğru sıralamada grubun 95'inci yüzdeliğinde ise aşağıdakilerden hangisi daima doğrudur?

- A) Öğrencinin notu A olacaktır.
- B) Öğrenci en azından alınabilecek bütün puanların %95'ini kazanmıştır.
- C) Öğrenci grubun %95 ini geride bırakmıştır.
- D) Öğrenci grubun %5'ini geride bırakmıştır.

Çünkü

- 1) Yüzdeler küçükten büyüğe sıralanmış verilerin belli bir yüzdesini altında bırakan noktadaki değerdir.
- 2) 95'inci yüzdelik ifadesi ölçümlerin %95 ini üstünde %5 ini altında bulunduran noktadaki değerdir.
- 3) 95'inci yüzdelik en fazla 100 alınabilecek bir sınavdan 95 alan bireylerin yerinin gösterildiği noktadır.
- 4) 95'inci yüzdelik grubun baştan 5. sırasındaki bireydir.

3. Bir sınıftaki öğrencilerin notları 83, 88, 90, 88, 86, 84 tür. Bu sınıfta 90 alan öğrencinin puanı arttırılırsa,

- A) Medyan değişir, standart sapma değişmez.
- B) Medyan değişmez, standart sapma değişir.
- C) Medyan da standart sapma da değişir.
- D) Bir tek ranj değişmez.

Çünkü

- 1) Bir seride medyanı ve standart sapmayı değiştirmeden serideki sayılar arttırılamaz.
- 2) Bir seride ki en yüksek puanın değişiminden bir tek ranj etkilenmez.
- 3) Bir seride ki en yüksek puanın değişimi medyanı etkilemez.
- 4) Bir seride ki en yüksek puanın değişimi standart sapmayı etkilemez.



4. Aşağıda sayıların bir kümesi verilmiştir. Bu dağılımın hangi merkezi eğilim ölçüsü ile temsil edilmesi en uygundur?

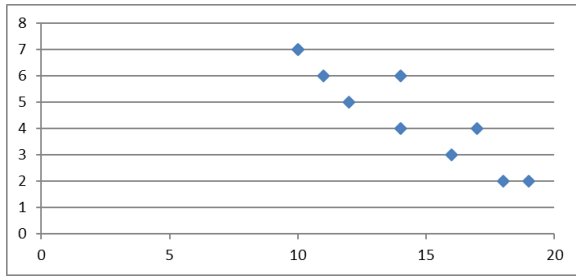
3, 4, 5, 6, 6, 8, 10, 12, 19, 26, 83

- A) Aritmetik Ortalama
- B) Ortanca
- C) Tepe değeri
- D) Geometrik Ortalama

Çünkü

- 1) 83 uç değerini de dâhil edebileceğimiz bir istatistik seçmeliyiz.
- 2) Bütün değerlerin kullanılarak hesaplandığı bir istatistik seçmeliyiz.
- 3) En fazla frekansın gözleendiği değerler bize grubu daha iyi temsil etme şansı verir.
- 4) Veri kümesinin yanlış değerlendirilmesine neden olacak uç değerlerin etkisini azaltmalıyız.

5.



Yukarıda X ve Y değişkenleri için verilen dağılım grafiğinden yararlanarak X ve Y arasındaki korelasyon katsayısını ..... olarak tahmin edebiliriz.

- A) -0.1
- B) 0.0
- C) 0.8
- D) -0.9

Çünkü

- 1) Veriler arasında ilişki yoktur.
- 2) Veriler arasında güçlü bir ilişki vardır.
- 3) Verilerden biri artarken diğeri de artmaktadır.
- 4) Veriler arasında zayıf bir ilişki vardır.

6. Aşağıdaki istatistiklerden hangisinin normal dağılım göstermesini beklersiniz?

- A) İlköğretim 5. Sınıf öğrencilerinin yaşları
- B) İlköğretim 3. Sınıf öğrencilerin zekâ düzeyleri
- C) A şehrindeki insanların cinsiyetleri
- D) 30 kişiyi aşan grupların özellikleri

Çünkü bir veri grubunun normal dağılım göstermesi;

- 1) Frekans değerlerinin bir değer etrafında toplanmasına bağlıdır.
- 2) Her bir alt grubun frekans değerlerinin eşit olmasına bağlıdır.
- 3) Frekans eğrisinin çan şeklinde ve simetrik olmasına bağlıdır.
- 4) Gruptaki kişilerin sayıca 30'u aşmasına bağlıdır.

7. Bir bilim insanı yaptığı deney sonunda 50 ayrı ölçme sonucu elde ediyor. Bu verilerin standart sapması -0,8 olduğuna göre aşağıdakilerden hangisi kesinlikle doğrudur?

- A) Ölçümlerin çoğu negatiftir
- B) Ölçümlerin tümü ortalamadan daha küçüktür.
- C) Standart sapma yanlış hesaplanmıştır.



D) Ölçümlerin tümü negatiftir.

Çünkü

- 1) Standart sapma ortalamadan farkların toplamıdır.
- 2) Standart sapma -1 ile +1 arasında değerler alabilir.
- 3) Standart sapma ölçümler negatif iken negatif değer alır.
- 4) Standart sapma negatif değer alamaz.

**8.** Masanın uzunluğunu ölçmek için kullanılan ölçme aracı metre ise bu ölçme aracının birimleri hakkında ne söylenebilir?

- A) Birimi yapaydır.
- B) Birimi kullanışlı değildir.
- C) Birimi doğaldır.
- D) Birimleri eşit değildir.

Çünkü

- 1) Ölçme aracının sıfır noktası mutlak sıfırdır.
- 2) Ölçme aracı birimleri eşit olarak ayrılır.
- 3) Ölçme aracı birimleri ortak kararlarla belirlenir.
- 4) Ölçme aracı ölçmek istediği özelliği direk olarak ölçer.

**9.** Ali öğretmen, her sorunun 10 puan olmasına karar verdiği sınavı için bir uygulama yapmış. Uygulama sonunda sınav kâğıtlarına 0 ile 100 arasında puanlar vermiştir. Ali öğretmenin yaptığı işlem nedir?

- A) Ölçmedir.
- B) Değerlendirmedir.
- C) Ölçüttür.
- D) Ölçümdür.

Çünkü

- 1) Ölçme sonuçlarını bir ölçütle karşılaştırmıştır.
- 2) Ölçme kuralını belirlemiştir.
- 3) Gözlem sonucunu sayılarla ifade etmiştir.
- 4) Ölçümleri ölçütlerle karşılaştırarak karara varmıştır.

**10.** Ayşe öğretmenin yaptığı sınavda öğrencileri düşük puanlar almıştır. Bu nedenle Ayşe öğretmenin her öğrenciye almış olduğu puanın %10 fazlasını vermiştir. Ayşe öğretmenin yaptığı işlem nedir?

- A) Ölçme sonucu ham puan elde etmektir.
- B) Ölçme sonuçlarına sabit hata karıştırmaktır.
- C) Ölçme sonuçlarına sistematik hata karıştırmaktır.
- D) Ölçme sonuçlarına tesadüfî hata karıştırmaktır.

Çünkü

- 1) Hatanın tespit edilmesi zorlaşmıştır.
- 2) Bütün ölçme sonuçları aynı miktarda etkilenmiştir.
- 3) Öğrencilerin başarısızlıklarından kaynaklanmıştır.
- 4) Ölçme sonuçları belli oranlarda değişmiştir.

**11.** Aşağıda güvenilirlik ve geçerlik ile ilgili olarak verilen karşılaştırmalardan hangisi doğrudur?

- A) Geçerli her test güvenilirlidir.
- B) Geçerlik ve güvenilirlik birbirini etkileyen kavramlar değildir.

- C) Güvenilir her test geçerlidir.  
D) Geçerli her test güvenilir, güvenilir her test de geçerlidir.

Çünkü

- 1) Bu özelliklerden biri sağlanıyor ise diğeri de sağlanmış olur.
- 2) Hatalardan arınık bir test ölçmek istediği özelliği ölçülebilir.
- 3) Testin ölçmek istediği davranışı ölçebilmesi için önce hatalardan arınması gerekir.
- 4) Bunlar bağımsız kavramlardır.

**12.** Tutarlılık, duyarlılık, kararlılık gibi kavramlar ölçme aracının hangi özelliği ile ilgilidir?

- A) Kullanışlılık  
B) Geçerlilik  
C) Güvenilirlik  
D) Objektiflik

Çünkü bu özellikler

- 1) Teste daha az hata karışığının göstergesidir.
- 2) Testin ölçme istediği davranışları ölçtüğünün göstergesidir.
- 3) Testin çok yorulmadan kullanılabilmesinin göstergesidir.
- 4) Puanlamaya puanlayıcı hatasının karışmamasıdır.

**13.** Aşağıda A ve B gruplarına uygulanmış iki ayrı testin istatistikleri yer almaktadır.

A ve B grubuna uygulanmış iki ayrı testin istatistikleri			
	Puan Aralığı	Standart Sapması	Ortalaması
A Testi	0-100	10	50
B Testi	0-10	2	5

Gruplardan hangisi ölçülen özellik bakımından daha çok farklılık göstermiştir?

- A) A testini alan grup  
B) B testini alan grup  
C) İkisi de eşittir.  
D) Verilen istatistikler yeterli değildir.

Çünkü iki grubun farklılaşması karşılaştırılırken;

- 1) Standart sapma dikkate alınır.
- 2) Puan aralıklarının genişliği dikkate alınır.
- 3) Basıklık katsayısı verilmeden yorum yapılamaz.
- 4) Standart sapma ve ortalama birlikte dikkate alınır.

**14.** Bir testin 8. maddesinin güçlüğü 1'e çok yakınsa, bu maddenin ayırt ediciliği kaç olabilir?

- A) 0'a çok yakın  
B) -1'e çok yakın  
C) +1'e çok yakın  
D) Tahmin için uygun veri yoktur.

Çünkü

- 1) Maddeyi doğru cevaplayanların çok az olması nedeniyle zayıf olan öğrenciler daha iyi ayırt edilebilir.

- 2) Maddeyi doğru cevaplayanların çok olması, iyi ve iyi olmayan öğrencileri ayırt etmeyi zorlaştırır.
- 3) Maddeye neredeyse herkesin yanlış cevap vermiş olması, bilenle bilmeyeni ayırt etmeyi zorlaştırır.
- 4) Madde güçlüğünden elde edilen veri ile madde ayırt ediciliğini tahmin edilemez.

15.Yapılan bir çalışmada lise 1 öğrencilerinin eleştirel düşünme ve problem çözme becerileri arasındaki korelasyon - 0,85 çıkmıştır. Bu sonuçlara göre eleştirel düşünme ve problem çözme arasında nasıl bir ilişki vardır?

- A) Düşük bir ilişki vardır.
- B) Bir ilişki bulunamamıştır.
- C) Yüksek bir ilişki vardır.
- D) Korelasyon yanlış hesaplanmıştır.

Çünkü

- 1) Korelasyon eksi çıktığı için aralarında hiç ilişki yoktur sonucuna varabiliriz.
- 2) Korelasyon eksi çıktığı için aralarında zayıf bir ilişki olduğu sonucuna varabiliriz.
- 3) İlişkinin gücü yorumlanırken korelasyon katsayısının işareti göz önünde bulundurulmaz.
- 4) Korelasyon eksi çıkamayacağı için yanlış hesaplanmıştır diyebiliriz.

16. Bu soru için doğru cevap C ise en iyi çalışan çeldirici hangisidir?

- A) A
- B) B
- C) C
- D) D

Çünkü

- 1) Üst gruptan en çok kişiyi çeken şık en iyi çeldiricidir.
- 2) Toplamda en çok işaretlenen şık en iyi çeldiricidir.
- 3) Alt grup- üst grup farkının fazla olduğu seçenek en iyi çeldiricidir.
- 4) Üst grup- alt grup farkının fazla olduğu seçenek en iyi çeldiricidir.

17.

Soru	Güçlük indeksi	Ayırt edicilik gücü
1	0,75	0,80
2	0,25	0,20
3	0,75	0,20
4	0,25	0,80

Yukarıdaki istatistiklere göre en ayırt edici ve zor olan soru hangisidir?

- A) 1.soru
- B) 2.soru
- C) 3.soru
- D) 4.soru

Çünkü sorunun hem ayırt edici hem de zor olması için;

- 1) Ayırt edicilik gücü 1'e, güçlük indeksi 0'a yaklaşmalıdır.
- 2) Ayırt edicilik gücü 0'a, güçlük indeksi 1'e yaklaşmalıdır.
- 3) Ayırt edicilik gücü 1'e, güçlük indeksi 1'e yaklaşmalıdır.
- 4) Ayırt edicilik gücü 0'a, güçlük indeksi 0'a yaklaşmalıdır.