

## Angoff (1-0), Nedelsky ve Sınır Değerleri Saptama Yöntemleri ile Bir Testin Sınıflama Doğruluklarının İncelenmesi

### *Angoff (1-0), Nedelsky and Examination of Classification Accuracies of a Test by Determination Methods of Limit Values*

Fazilet TAŞDEMİR\*  
Ömay ÇOKLUK\*\*

**Özet:** Bu araştırma, sınıf öğretmenliği öğrencilerinin test başarılarının Angoff (1-0), Nedelsky standart belirleme yöntemleri ve sınır değer saptama yöntemleri ROC ve Aralık Tahmini ile belirlenen sınıflama doğruluklarının incelenmesini içermektedir. Araştırmada farklı standart belirleme yöntemleri incelenmiş olup araştırma, tarama modelinde korelasyonel bir araştırmadır. Araştırmada veri toplama aracı olarak, 2006-2012 yıllarında uygulanan KPSS'deki Türkçe ve Matematik sorularından tesadüfi olarak seçilmiş toplam 30 sorudan oluşan bir test kullanılmıştır. Araştırmanın sonunda, en yüksek kesme puanı Türkçe ve Matematik testi için Angoff (1-0) yöntemi ile elde edilmiş, başarılı sayılan öğrenci yüzdeleri arasında manidar farklılık olduğu sonucuna ulaşılmıştır. Türkçe ve Matematik testlerinin kesme puanlarının belirlenmesi için uygulanan ROC analizi sonucunda, Matematik testinin başarılı/başarısız olan öğrencileri doğru sınıflandırabildiği görülmüştür. Matematik ve Türkçe testi için aralık tahminine göre belirlenen kesme puanları, ROC analizinde belirlenen kesme puanlarla uyumlu bulunmuştur. Sonuç olarak, Türkçe ve Matematik alt testleri için Angoff (1-0) ve Nedelsky; ROC ve Aralık Tahmini yöntemlerine göre farklı kesme puanları belirlenmiştir.

**Anahtar sözcükler:** Angoff (1-0), Nedelsky, ROC, Kesme puan

**Abstract:** This research includes the examination of classification accuracies of test successes of class-teachership students by Angoff (1-0), Nedelsky standard determination methods, limit determination methods, ROC and interval forecast. In this research, different standard determination methods have been examined and the research is correlational in scanning model. In the research, a test consisting of 30 questions (selected coincidentally from Turkish and math questions in KPSS which was held in the years 2006/2012) was used as a data collection tool. At the end of this research, the highest cut-off scores for Turkish and mathematics tests were obtained by the Angoff (1-0) method, and it has been concluded that there exists a significant difference among the student percents that are deemed to be successful. As a result of the ROC analysis in order to determine the cut-off scores for Turkish and mathematics tests, it has been observed that the mathematics test is capable of classifying the successful/unsuccessful students accurately. It has shown that the cut-off scores determined according to interval estimation for mathematics and Turkish test conform to the cut-off scores determined in ROC analysis. As a result, Angoff (1-0) and Nedelsky for General Ability sub-tests and different cut-off scores for ROC and Interval Estimation Methods have been identified.

**Keywords:** Angoff (1-0), Nedelsky, ROC, Cut-off Score

\* Arş. Gör., Ankara Üniversitesi, EBF, Ölçme ve Değerlendirme Bölümü, Ankara, fazilettasdemir@yahoo.com

\*\* Doç. Dr., Ankara Üniversitesi, EBF, Ölçme ve Değerlendirme Bölümü, Ankara, cokluk@education.ankara.edu.tr

Eğitimde ve psikolojide kişi hakkında çeşitli kararlara dayanak oluşturmak için bilgi toplamının çeşitli yolları vardır. En sık tercih edilen yol, kişiyi bir test ile karşı karşıya getirmek ve kişinin testteki maddelere gösterdiği tepkileri belli bir yöntemle göre puanlamaktır. Testler diğer yöntemlere göre daha kolay uygulanabilmesi, objektif olması, geçerli ve güvenilir gözlemler yapmaya olanak vermesi sebebiyle tercih edilmektedir (Cronbach, 1990; Conway, 2006; Pawar, 2004) ve kişilerin çeşitli eğitsel, mesleki alanlardaki yeterlilikleri hakkında karar vermek amacıyla kullanılır. Yeterlilikleri konusunda karar vermek için kişilerin yanıtlarına bağlı olarak testten aldıkları puanlar bir ölçüt ile karşılaştırılır. Bu yolla danışma hizmetlerine, sınıflamaya, seçme ve yerleştirmeye yönelik kararlar alınabildiği gibi, öğretim programlarının ve uygulamalarının etkililiğine ilişkin kararlar da alınabilmektedir (Anastasi, 1988; Kubiszyn, & Borich, 2003). Değerlendirmede kararın doğruluğu hem ölçmenin, hem de ölçütün doğruluğuna bağlıdır. Sınavdan başarılı sayılabilmek için alınması gereken minimum puan amaca uygun biçimde belirlenmemiş ise, ölçme işlemi hatalardan arınmış olsa da ölçütün hatalı belirlenmesine bağlı olarak hatalı kararlar alınabilir (Özçelik, 1992; Turgut, 1992). Değerlendirmeler yaparken kişinin test puanını yorumlamada iki farklı ölçüt türü esas alınabilir. Kişinin test puanı, testi alan diğer kişiler içindeki puanları ile karşılaştırılıyorsa bu tür değerlendirmeler bağıl değerlendirme, kişinin test puanı önceden belirlenmiş bir davranış içeriği ile ilişkilendirilerek değerlendirme yapılyorsa, bu tür değerlendirmeler mutlak değerlendirme olarak adlandırılmaktadır. Mutlak değerlendirmede kişinin gruptaki diğer kişilere göre nasıl bir performans gösterdiği önemli değildir, önemli olan kişinin ne yapabildiğidir (Crocker, & Algina, 1986; Cronbach, 1990; Goodwin, 1996; Mathison, 2005). Performans, test maddelerine bağlı olarak ölçüldüğünde tanımlayıcı bilgiler elde edilmiş olur. Kişinin performans düzeyinin eğitim kurumlarının müfredatla hedeflediği amaçlar için yeterli olup olmadığı gibi soruları cevaplayabilmek için standartlar belirlenmelidir. Böylece önceden standardı belirlenmiş testlerin kullanımının sağlanması ile birlikte, kişilerin ne yapabildiği konusunda bilgi sağlanabilir.

### 1) Standart Belirleme

Standart belirleme (standard setting) kavramı, alan yazında test maddelerinin doğru yanıtlanma olasılıklarından yola çıkarak bir testten alınacak puanları, belirlenen kesme puanları ile belli düzeylere ayırma ve bu düzeylerin ne anlama geldiğini tanımlama anlamında kullanılmaktadır (Shen, 2001; Zwick, Şentürk, & Wang, 2001; Schulz, 2006). Farklı başarı düzeylerine ilişkin bir sınıflama yapabilmek için her düzeye ilişkin kesme puanlarının ortaya konulması gerekmektedir. Standart belirleme süreci kişilerin başarı düzeylerini ayırıştırabilmek için son derece önemlidir. Bu açıdan kullanılacak olan standart belirleme yönteminin seçimi de önem taşımaktadır (Cizek, 2001). Bu bağlamda standart belirleme yöntemlerinden söz etmekte yarar vardır.

### 2) Standart Belirleme Yöntemleri

Jaeger (1989) Meskuaskas'ın süreç yöntemleri adını verdiği bu yöntemleri "*Test Merkezli Yöntemler*" ve "*Öğrenci Merkezli Yöntemler*" olarak ikiye ayırmıştır. Test Merkezli Yöntem'de yargıcı kararları testin içeriğine göre verilir. Test maddeleri ya ayrı ayrı incelenir ya da testin tamamı bütünsel anlamda ele alınır. Öğrenci Merkezli Yöntemlerde ise ele alınan konu ile ilgili beklenen yeterlik düzeylerine göre öğrencilerin yeterlilikleri hakkında kararlar alınır. Standart belirleme yöntemleri, Jaeger'ın yapmış olduğu Test Merkezli ve Öğrenci Merkezli Yöntemler sınıflaması temelinde aşağıda sunulmaktadır.

#### 2.1. Öğrenci Merkezli Yöntemler

Öğrenci Merkezli Yöntemlerin en bilinenleri Zıt Gruplar Yöntemi ve Sınır Grup Yöntemi'dir. Bu yöntemlerde kesme puanlarının nasıl belirlendiği, yöntemler arası benzerlik ve farklılıklar

aşağıda açıklanmaktadır.

**2.1.1. Zıt Gruplar Yöntemi:** Yargıcılar tarafından “yeterli” ve “yetersiz” olduğu düşünülen kişilerden oluşan iki zıt gruba test uygulanır. Her bir puan düzeyindeki öğrencinin “yeterli” grupta yer alma olasılığı hesaplanır. Sınır düzeyde tanımlanan kişiye ilişkin bilgi kaynaklarından öncelikli olarak faydalanılmaktadır (Cizek, & Bunch, 2007).

**2.1.2. Sınır Grup Yöntemi:** Değerlendirilecek kişilerin yeterli ya da yetersiz olarak iki sınıfa ayrıştırılması kolay gibi gözükse de, bazı yargıcılar için kolay olmayabilir. Yargıcıların bazı kişileri ne tam yeterli, ne de tam yetersiz olarak tanımlayamadığı durumlar olabilir. Livingston ve Zeiky (1989) yeterli ya da yetersiz sınıflandırmasının net olarak yapılamamasına dayanarak sınır grup yöntemini ortaya çıkarmışlardır. Bu yöntemde yargıcılardan yeterli ya da yetersiz olarak iki sınıflandırma yapmaları istenmektedir.

## 2.2. Test Merkezli Yöntemler

Test Merkezli Yöntemler, maddeler ile ilgili yargıcıların vermiş oldukları kararlarına dayanır. Test Merkezli Yöntemler Nedelsky Yöntemi, Ebel Yöntemi, Jeager Yöntemi ve Angoff Yöntemi ile Angoff Yöntemi'nin farklılaştırılmasından oluşan Angoff (1-0) çeşitlemesi olarak bilinmektedir. Test Merkezli Yöntemler aşağıda açıklanmıştır:

**2.2.1. Nedelsky Yöntemi:** Bu yöntem yalnızca çoktan seçmeli testlerde kullanılabilir; çünkü yöntem, “*öğrenci testi çözerken seçenekleri eleme sürecini izler, yanlış olduğunu düşündüğü seçenekleri eleyerek doğru cevaba ulaşır*” fikrini temel almaktadır (Livingston, & Zeiky, 1989). Nedelsky, bu yöntemde geçme-kalma sınırındaki ya da minimum yeterlik düzeyindeki bir öğrencinin sorunun yanlış seçeneklerini eledikten sonra kalan seçenekler arasından rastgele bir tercih yaparak soruyu cevapladığını belirtir. Yöntem, yargıcıların test içerisindeki her bir maddeyi dikkatle inceleyerek “*minimum yeterlik düzeyindeki öğrencinin*” (D-F yani geçme-kalma sınırındaki öğrencinin) yanlış cevabı eledikten sonra kalan seçenekler arasından doğru cevap bulma olasılığını hesaplamaya çalışır. Her bir soruya ilişkin her bir yargıcının yaptığı tahminler toplamının ortalamaları alınarak, o soruya ilişkin Nedelsky değeri elde edilir. Minimum yeterlik düzeyindeki öğrencinin teste ilişkin puanını hesaplamak için her bir soruya ilişkin oluşturulan Nedelsky değerleri toplanarak toplam puan hesaplanır (Nedelsky, 1954).

**2.2.2. Ebel Yöntemi:** Ebel yönteminde yargıcıların test içerisinde yer alan her bir soruyu dikkatle inceleyerek iki boyutta değerlendirme yapması gerekir. Sorunların zorluğu ve önemliliğine ilişkin olan bu boyutlar kendi içerisinde sırasıyla üç ve dört gruba ayrılmaktadır. Yargıcıların her bir soru için 12 ayrı hücre temelinde inceleme yapması gereklidir. Her bir hücre için yargıcı, minimum yeterlik düzeyindeki (geçme-kalma sınırındaki) kişinin doğru cevaplama olasılığını hesaplamaya çalışır (Cizek, & Bunch, 2007).

**2.2.3. Jeager Yöntemi:** Madde temelinde değerlendirme yapmayı gerektiren bir diğer yöntemdir. Angoff yöntemi ile benzer özellikler taşımaktadır. Bu yöntemde de yargıcılar her bir soruyu, “*değerlendirilen herkes bu soruyu doğru cevaplayabilir mi?*” sorusunu dikkate alarak değerlendirir (Cizek, 2001). Yöntemin Nedelsky yönteminde olduğu gibi, elde edilecek olasılık değerinin 0.50 ile 1.00 arasında bir değer almasına dayalı bir sınırlılığı söz konusudur.

**2.2.4. Angoff Yöntemi:** Eğitim araştırmalarında en çok bilinen ve kullanılan standart belirleme yöntemidir. Bu yöntemde uzmanlardan oluşan yargıcılardan minimum yeterlik düzeyindeki öğrencilerin testte bulunan her maddeyi cevaplayabilme durumu hakkında yargıya varmaları istenir (Irwin, 2007). Her yargıcının testte bulunan bütün maddelere ilişkin sonuçları toplanarak o yargıcıya ait “*Minimum Geçme Puanı*” (MGP) hesaplanır. Angoff yöntemi diğer yöntemlere göre hem uygulanması hem de yargıcılara puanlamanın nasıl yapılacağına açıklanması açısından

dan daha kolay bir yöntemdir (Berk, 1986).

**2.2.5. Angoff Çeşitlemesi (1- 0) Yöntemi.** Impara ve Plake (1997) tarafından geliştirilen bu yöntem Angoff çeşitlemesi olarak adlandırılmaktadır. Angoff yönteminde yargıcılardan, öğrenciye uygulanacak olan testteki her bir maddeyi inceleyip, minimum yeterlilik düzeyindeki bir öğrencinin bu maddeleri doğru cevaplayıp cevaplayamayacağını tahmin etmeleri istenir. Öğrencinin maddeyi doğru cevaplayacağı düşünülüyorsa “1” puan, yanlış cevaplayacağı düşünülüyorsa “0” puan verilir. Puanların toplamları alınarak o yargıcı için MGP belirlenir; tüm yargıcıların MGP ortalaması ile de testin kesme puanı belirlenmiş olur. Geleneksel Angoff yöntemi ile yapılan yüzde tahminlerinde yargıcılar verdikleri oranlardan pek fazla emin olamazken, Angoff (1-0) yönteminde iki boyuta düşürülen tahminlerin geçerliliğinin daha yüksek olduğu belirlenmiştir.

Standart belirleme yöntemleri ile kesme puanı belirleme dışındaki diğer kesme puanı belirleme yöntemleri de sınır değer saptamaya dayalı yöntemlerdir. Bu yöntemler de Test Merkezli standart belirleme yöntemlerinde olduğu gibi, testlerde kesme puanı olarak hangi değer alınacağına karar vermek ve kesme puanına göre sınıflama doğruluklarını belirlemek amacıyla günümüzde Tıp, Veterinerlik, Sağlık Bilimleri gibi alanlarda kullanılmaktadır. Sınır değer saptamaya dayalı kesme puanı belirleme yöntemleri olan İşlem Karakteristik Eğrisi (Receiver Operating Characteristics Curve-ROC) ve Aralık Tahmini Yöntemleri sırasıyla aşağıda açıklanmıştır.

### 3) Sınır Değer Saptamaya Dayalı Standart Belirleme

Bir testin performansı, olguları doğru olarak gruplara (başarılı/başarısız) ayırabilme kapasitesi ile tanımlanabilir (Campell, & Zweig, 1993; Taga, Aslan, & Güner, 2000). Başarılı ve başarısız bireylerden oluşan iki grup arasında ayırım yapmak için her olası kesim noktası değerlendirilir. İyi bir testin yüksek duyarlılık, özgüllük, doğruluk ve yüksek pozitif ve negatif tahmin değerine sahip olması beklenir. Bir testin alt-üst sınırlarının belirlenmesi için örneklemden seçkisiz olarak seçilmiş kişilerin sayısı evreni temsil edecek büyüklükte olmalıdır. Test sonuçlarının betimleyici istatistikleri hesaplanır. Bu işlemin ardından kesme puanlara göre belirlenen grupların sınır değer belirleme yöntemlerinin kullanımına uygun olup olmadığına bakılır. Örnekleme yer alan kişilerin seçimi belirlendikten sonra dağılım sınanarak normal dağılıma uymayan gruplarda normal dağılımı elde edebilmek için istatistiksel yöntemlerle dönüşüm yapılmalıdır (Solberg, 1983). Bu aşamadan sonra kesme puanların belirlenmesinde kullanılacak sınır değer saptama yöntemleri uygulanır. Shapiro (1999), bu yöntemleri aşağıdaki sırayla şöyle açıklanmaktadır:

#### Yöntem 1. İşlem Karakteristik Eğrisi (ROC) ile Kesme Puanı Belirleme

İşlem Karakteristik Eğrisi (Receiver Operating Characteristics Curve-ROC) istatistik karar teorisine dayanır. ROC yöntemi, bir testin optimal duyarlılığını ve optimal özgüllüğünü belirlemek için uygun kesim noktalarının belirlenmesini sağlar. ROC eğrisi yardımı ile uygun kesme puanı belirlemek için bir referansa gereksinim vardır. ROC eğrisi, kişilerin testlerde gösterdikleri performansları belirlemek ve lojistik modeller, doğrusal sınıflama analizi gibi istatistiksel modellerin doğruluğunu değerlendirmek amacı ile kullanılan bir yöntemdir. Aynı zamanda testlerin ve tahmin modellerinin değerlendirilmesi, doğruluklarının sayısal sonuçlar ile açıklanması ayrıca tahminlerin doğruluğu arasında karşılaştırma yapılmasına da olanak verir (Obuchowski, Lieber, & Wians, 2004). Bir ROC eğrisi, farklı eşik değerleri için dikey eksen üzerinde doğru pozitiflik (duyarlılık) ve yatay eksen üzerinde yanlış pozitiflik (1-özgüllük) oranlarının yer aldığı bir eğridir. Kesme noktası olarak hangi değer alınacağını belirlemek için analiz sonucunda verilen her bir duyarlılık ve 1- özgüllük değerleri incelenir ve optimum nokta seçilir. Bu noktalar birleş-

tilerle bir içbükey eğri elde edilir. Bu eğri altında kalan alan testin ayırma (sınıflama) gücünü verir (Dirican, 1991; Knapp, & Miller, 1992; Wagner, 2007). Yanlış değerlere sahip olmayan ideal bir testte ROC eğrisi (0,0), (0,1) ve (1,1) noktalarını birleştirmektedir. Performansı düşük olan ROC eğrisi (0,0)'dan (1,1)'e kadar 450 açı yaparak uzanan köşegen şeklindedir. ROC eğrisi genellikle iki durum arasında değişir. Testin sonucu ne kadar iyi ise eğri o kadar yukarıya yani yüksek duyarlılık bölgesine ve sola, düşük yanlış pozitif oranı bölgesine doğru kayar. Buna karşın ROC çizimi  $y=x$  fonksiyonuna yaklaştıkça test sonucuna göre başarısı düşük olan bir eğri ortaya çıkar. Çünkü bu testte yanlış değerlerin oranı en yüksektir (Grove, 2006).

**ROC Eğrisi Altında Kalan Alan (Area Under Curve - AUC).** ROC eğrisi altında kalan alan (Area Under Curve - AUC), testin başarılı ve başarısız kişileri ayırmadaki doğruluk oranını belirler. Testin hiç ayırma yeteneği olmadığı durumda, ROC eğrisi altındaki alanın beklenen değeri 0.50'dir. Mükemmel bir testte ise sıfır yanlış pozitif ve sıfır yanlış negatif ile alanın değeri 1.00 olacaktır. Test, bu iki değer arasında bir alana sahip olmalıdır (Kanık ve Erden, 2003). ROC eğrisinin altında kalan alan ne kadar büyük ise, bu testin doğru sınıflama tahminlerinde o kadar iyi olduğunu gösterir (Hanley, & McNeil, 1982; Grove, 2006). ROC eğrisi altında kalan alanın (AUC) hesaplanması için farklı yöntemler kullanılmaktadır. Bu yöntemler aşağıda açıklanmaktadır:

## 1. Parametrik Yöntemler

### 1. a. AUC'nin Olasılık Formülleri ile Hesaplanması

Parametrik yöntem McClish (1989) tarafından geliştirilmiştir. Bu yöntemde, iki grup olduğu ve ölçüt değişken değerinin tüm kişiler için kullanılan bir özellik olduğu varsayılır. Başarılı grupta ölçüt değişkeni X ile başarısız gruptaki ölçüt değişkeni Y ile gösterilir ve olasılık dağılışı ifadeleri Formül 1'deki gibi gösterilir. Parametrik modelde X (başarılı kişiler) ve Y (başarısız kişiler) değişkenlerinin farklı ortalama ve varyansa sahip ayrı iki normal dağılım gösterdiği varsayılır (Karayianni, Tretiak, & Herrmann, 1996; Metz, & Pan, 1999).

$$X \approx N(\mu_x, \sigma_x^2), Y \approx N(\mu_y, \sigma_y^2)$$

(Formül 1)

ROC eğrisi, Formül 2'deki eşitlikten yararlanılarak çizilir (Karayianni, Tretiak, & Herrmann, 1996; Metz, Herman, & Shen, 1998).

$$YPO(c) = \Phi\left[\frac{\mu_x - c}{\sigma_x}\right], DPO(c) = \Phi\left[\frac{\mu_y - c}{\sigma_y}\right], -\infty < c < \infty$$

(Formül 2)

Formülde bulunan c, ölçüt değeridir. Doğru ve yanlış pozitiflik olasılık fonksiyonlarının birleşik fonksiyonundan AUC hesaplanabilir. Bu yöntemle çizilen ROC eğrisi düzgün değişen eğridir (McClish, 1989; Karayianni, Tretiak, & Herrmann, 1996; Metz, Herman, & Shen, 1998).

### 1.b. AUC'nin Çift Yönlü Olasılıklı Grafik Kağıdı ile Hesaplanması:

Bu yöntemde, her iki eksenin olasılık işaretli grafik kağıdı üzerinde binormal regresyon hattının eğim ve kesim noktalarını kullanarak alan hesabı yapılabilir (Hanley, & McNeil, 1982):

$$A = \Phi\left[\frac{a}{\sqrt{1+b^2}}\right]$$

(Formül 3)

Formül 3’de gösterilen “a”, ROC eğrisinin kesim noktasını ve “b”, ROC eğrisinin eğimini ifade eder (Hanley, & McNeil, 1982).

## 2. Nonparametrik Yöntem ile AUC Hesaplanması

Nonparametrik yöntem, AUC hesaplanmasında kullanılan bir yöntemdir (Hanley, & McNeil, 1982, 1983; DeLong, DeLong, & Clarke-Pearson, 1988). Parametrik yöntemin kullanılabilmesi için gerekli olan güçlü normallik varsayımlarının karşılamadığı durumlarda bu yöntem de yaygın olarak kullanılmaktadır (DeLong *et alii*, 1988). Seçilen başarılı ve başarısız grup eşleştirildiğinde bu iki grup değerlerindeki benzeşim, bunların doğru sınıflandırılmalarına yardımcı olur. Seçilen başarılı ve başarısız grupların doğru olarak sıralanma olasılığı Wilcoxonistatistiği ile ölçülür. Daha sonra da W ( $\theta$  olasılığı) hesaplanabilir (Hanley, & McNeil, 1982; DeLong *et alii*, 1988).

### Yöntem 2. Aralık Tahmini ile Standart Belirleme

Bu yöntemde test puanı değişkeninin, başarılı ve başarısız gruplarda normal dağılıma uygun olup olmadığı Kolmogorov-Smirnov tek örneklem testi ile sınanır. Sınama sonucunda başarılı ve başarısız grubun normal dağılıma uygunluk gösterdiği belirlenirse başarılı (kontrol) grupta teste ait “*aritmetik ortalama + (1,96) x standart sapma*” formülü yardımıyla testin üst sınırı belirlenir. Başarısız grubun test puanları ile de “*aritmetik ortalama - (1,96) x standart sapma*” ifadesi kullanılarak testin alt sınırı saptanır. Başarısız grubun normal dağılıma uygunluk göstermediği görüldüğünde logaritmik dönüşüm yapılması gerekir. Daha sonra başarılı grubun “*log*” değişkenine ait “*aritmetik ortalama + (1,96) x standart sapma*” ve başarısız gruba ait “*aritmetik ortalama - (1,96) x standart sapma*” formülleri ile kesim değerlerinin bulunduğu aralık belirlenir. Bu sınırlarda test değerlerine ait duyarlılık ve özgüllükler hesaplanır (Shapiro, 1999).

### Yöntem 3. Yüzde Değerlerine Göre Standart Belirleme

Sınır değerlerin belirlenmesinde kullanılan bu yöntemde, testin başarılı ve başarısız grup sonuçlarının ortanca değeri dikkate alınarak kesme puanı hesaplanır (Letner, 1992). Başarılı grubunun maksimum değerine yakın değerler içinden belirlenen kesim değeri üst sınır; başarısız grubunun minimum değerine yakın değerler arasından belirlenen kesim değeri ise alt sınırı oluşturur. Bulunan bu kesim değerleri arasında kesim noktası için duyarlılık ve özgüllükler hesaplanır ve kesme puanı belirlenir (Shapiro, 1999).

Sınır değer saptama yöntemleri ile sağlık bilimlerinde yapılan araştırmalarda bireylerin sağlık durumlarıyla ilgili sınıflandırmalar yapmak amacıyla söz edilen yöntemlerden bir veya birkaçı kullanılmaktadır. Eğitim Bilimleri araştırmalarında ise bu amaçla yapılmış bir araştırma yer almamaktadır. Bu araştırmada, uygulama süreçlerinin Öğrenci Merkezli Yöntemlere göre daha az zaman alması ve kesme puanlarının hesaplanmasının daha pratik olması nedeni ile Test Merkezli Yöntemlerden Nedelsky ve Angoff (1-0) Yöntemleri ve bunun yanı sıra sınır değer saptama yöntemlerinden ROC ve Aralık Tahmini ile Standart Belirleme Yöntemleri ele alınıp incelenmiştir.

Araştırmalar incelendiğinde, Standart Belirleme Yöntemlerine göre elde edilen kesme puanları arasındaki farklılıklar ile ilgili bazı araştırmaların olduğu görülmektedir. Bunun yanı sıra yargıcılar arasındaki uyumun incelendiği araştırmalara da rastlanmaktadır. Yurt içi ve yurt dışında yapılan araştırmalarda Taşdelen (2009), Nedelsky ve Angoff Standart Belirleme Yöntemleri ile elde edilen kesme puanlarını karşılaştırmış, Ömür, & Selvi (2010) Angoff, Ebel, & Nedelsky yöntemleri ile farklı yargıcı sayıları kullanarak sınıflama tutarlılıklarını araştırmış; Hurtz ve Hertz (1999) Angoff yönteminde kaç yargıcı kullanılması gerektiğini ve Hess, Subhiyah,

& Giordano (2007) Kümeleme Analizi ile Angoff (1-0), Ebel ve Nedelsky Yöntemleri'ni kullanarak sınıflama doğrulukları ile ilgili kararları incelemiştir. Standart Belirleme Yöntemleri'nin birbirlerine göre avantajları ve dezavantajları, kesme puanı belirlemek için yargıcı yeter sayısı gibi problemler üzerinde durulmuştur. Bir diğer kesme puanı belirleme yöntemi olan sınır değer saptama yöntemlerinin ise sadece Tıp ve Sağlık Bilimlerinde hasta olup olmama durumunun sınıflandırma doğruluklarının belirlenmesinde sıklıkla kullanıldığı görülmektedir. İncelenen araştırmalarda birçok yöntemin bir arada kullanıldığı fakat Test Merkezli Yöntemlere dayalı olarak yargıcıların belirlediği kesme puanı belirleme yöntemleri ve sınır değerleri saptama yöntemleri ile belirlenen sınıflama doğruluklarının incelendiği ve kesme puanlarının isabetliliğinin sınıflandırmaya rastlanamamıştır. Bu nedenle kesme puanı belirlemede kullanılan farklı yöntemlerin incelenmesi ve bunları temel alan araştırmaların yapılması test sonuçlarına göre başarılı/başarısız şeklinde sınıflandırılan kişilerin doğru değerlendirilmesine de olanak sağlayacaktır.

Yukarıda açıklanan gerekçe ve sorunlardan dolayı bu araştırmanın problemini, bir testin standart belirleme yöntemleri Angoff (1-0), Nedelsky ve Sınır Değerleri Saptama Yöntemleri İşlem Karakteristik Eğrisi (ROC) ve Aralık Tahmini yöntemleri ile belirlenen sınıflandırma doğruluklarının incelenmesi oluşturmaktadır.

### **Amaç**

Bu araştırmanın genel amacı, sınıf öğretmenliği programı öğrencilerinin Türkçe ve Matematik sorularından oluşan bir teste göre belirlenen başarılı/başarısız sınıflandırılma durumlarının Test Merkezli standart belirleme ve sınır değer saptama yöntemleri kullanılarak incelenmesidir. Bu genel amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır;

- 1) Angoff (1-0) ve Nedelsky Yöntemleri ile belirlenen kesme puanlarına göre başarılı/başarısız olarak kabul edilen öğrenci yüzdeleri arasında manidar bir farklılık var mıdır?
- 2) Bir güvenilirlik kanıtı olarak, kesme puanlarının belirlenmesinde;
  - a. Angoff (1-0) Yöntemi kullanıldığında yargıcılar arası uyum katsayısı nedir?
  - b. Nedelsky yöntemi kullanıldığında yargıcılar arası uyum katsayısı nedir?
- 3) Angoff (1-0) ve Nedelsky Yöntemleri arasında, yargıcılardan elde edilen minimum geçme puanları (MGP) bakımından uyum var mıdır?
- 4) Sınır Değer Saptama Yöntemleri olan ROC ve aralık tahmini ile belirlenen kesme puanları nelerdir ve bu puanlar arasında uyum var mıdır?

### **Yöntem**

#### **Araştırmanın Modeli**

Bu araştırmada farklı standart belirleme yöntemleri incelenmiş olup araştırma, ilişkisel tarama modelinde korelasyonel bir araştırmadır. Korelasyonel araştırmalar, iki ya da daha çok değişken arasındaki ilişkinin, herhangi bir biçimde bu değişkenlere müdahale edilmeden incelendiği araştırmalardır. Korelasyonel araştırmaların, değişkenler arasındaki ilişkilerin açığa çıkarılması ve bu ilişkilerle ilgili daha üst düzey araştırmaların yapılması için gerekli ipuçlarını sağlayan önemli araştırmalar oldukları ifade edebilir (Büyüköztürk *et alii*, 2010). Araştırma, KPSS Genel Yetenek Testi'nin Angoff (1-0), Nedelsky ve Sınır Değer Saptama Yöntemlerine dayalı olarak kesme puanlarının belirlenmesi ve bu yöntemlerle belirlenen sınıflandırma doğruluklarının incelenmiş olması sebebiyle de temel bir araştırmadır. Temel araştırmaların amacı, var olan bilgiye yenilerini katmaktır (Karasar, 2005).

### Araştırma Grubu

Araştırma grubu belirlenirken seçkisiz olmayan örnekleme yöntemlerinden amaçsal örnekleme yöntemlerinden biri olan ölçüt örnekleme yöntemi kullanılmıştır. Amaçsal örnekleme araştırmanın amacına bağlı olarak bilgi açısından zengin durumların seçilip derinlemesine araştırma yapılmasına olanak sağlar. Bu çalışmada amaçsal örnekleme türlerinden kullanılan ölçüt örnekleme yönteminde, gözlem birimleri belli niteliklere sahip kişiler, olaylar, nesnelere ya da durumlardan oluşturulabilir (Büyüköztürk *et alii*, 2010). Araştırma grubu, ifade edilen yöntemde belirtildiği gibi bilgi açısından zengin durumların seçilmesi ve araştırma yapılmasına olanak sağlaması amacıyla Ankara İli devlet üniversitelerinden Ankara Üniversitesi Eğitim Bilimleri ve Hacettepe Üniversitesi Eğitim Fakülteleri'nde öğrenim görmekte olan Sınıf Öğretmenliği Programı'nın üçüncü sınıfındaki 211 öğrenciden oluşmaktadır. Bu çalışmada amaçsal örnekleme türlerinden ölçüt örnekleme yöntemi kullanılmasının sebebi ise, ölçüt örnekleme yönteminde gözlem birimlerinin belli niteliklere sahip kişiler, olaylar, nesnelere ya da durumlardan oluşabilmesidir. Bu nedenle sınıf öğretmenliği programının üçüncü sınıfındaki öğrenciler belirlenirken, 2008-2009 güz döneminde üniversiteye yerleştirilmiş ve Türkçe, Matematik dersi almış olan öğrenciler dikkate alınmıştır. Araştırmanın amacı doğrultusunda Angoff (1-0) ve Nedelsky standart belirleme yöntemlerine uygun biçimde kesme puanı belirlemek için yargıcı görüşü olarak, Türkçe ve Matematik alanındaki yargıcıların kararlarına başvurulmuştur. Bu çalışmada yargıcı görüşleri alınırken her bir yöntem için 10 farklı yargıcı olmak üzere toplam 40 yargıcıdan görüş alınmıştır.

### Veriler ve Toplanması

Araştırmada veri toplama aracı olarak, 2006-2012 yıllarında uygulanan KPSS Testi'ndeki Türkçe ve Matematik sorularından seçkisiz olarak belirlenmiş toplam 30 sorudan oluşan bir test kullanılmıştır. Türkçe ve Matematik soruları, asıl uygulama öncesinde üç alan uzmanının görüşüne sunulmuş ve hangi sorunun hangi konu alanı ile ilgili olduğunun belirlenmesine yönelik bir çalışma yapılmış ve soruların testteki yaklaşık ağırlıkları da dikkate alınarak KPSS Testi'ndeki Türkçe ve Matematik sorularından oluşturulan 30 soruluk bir test hazırlanmıştır. Araştırmada, Sınıf Öğretmenliği Programı öğrencilerinden test aracılığı ile alınan puanlar öğrenci verilerini oluşturmuştur. Hazırlanan genel yetenek testindeki sorular için Angoff (1-0) ve Nedelsky Yöntemleri ile kesme puanı belirlenirken seçkisiz olarak her iki yöntem için 20 yargıcıdan oluşan toplam 40 kişilik bir yargıcı grubunun görüşlerine başvurulmuştur. Yargıcı kararına dayalı veriler, araştırmacı tarafından hazırlanan içerisinde madde kartlarının bulunduğu "*Yargıcı Kararı Edinme Formu*" ile elde edilmiştir. Sınır değer saptama yöntemlerinin analizinde kullanılması için gerekli olan öğrencilerin lisans Türkçe ve Matematik dersleri akademik başarı not ortalamaları ile ilgili bilgiler, fakültelerin öğrenci işlerinden temin edilmiştir.

### Verilerin Analizi

Araştırmanın birinci alt amacı olan, Angoff (1-0) ve Nedelsky yöntemleri ile belirlenen kesme puanlarına göre başarılı/başarısız olarak kabul edilen öğrenci yüzdeleri arasında manidar bir farklılık olup olmadığının belirlenmesinde, bağımlı iki oran arasındaki farkın testi yapılmıştır. Angoff (1-0) ve Nedelsky yöntemi için alınmış olan yargıcı kararları ile her bir yargıcıya ait MGP belirlenmiştir. Kesme puanlarına göre her bir yöntem için başarılı sayılan öğrenci sayıları ve yüzdeleri belirlenmiştir. Yöntemler Angoff (1-0) ve Nedelsky olmak üzere karşılaştırılarak Z istatistiği hesaplanmıştır.

Araştırmanın ikinci alt amacı bir güvenilirlik kanıtı olarak, kesme puanlarının belirlenmesinde Angoff (1-0) Yöntemi ve Nedelsky Yöntemi kullanıldığında yargıcılar arası uyum



katsayısının ne olduğuna ilişkindir. Türkçe ve Matematik testlerinde Angoff (1-0) yöntemi için karar vermiş olan yargıcıların, her bir test maddesi için verdikleri kararlarının uyumuna Cochran Q Testi ile bakılmıştır. Nedelsky Yöntemi'nde yargıcılar arası uyum katsayısının belirlenmesinde ise Sınıf (Küme) içi Korelasyon Katsayısı (Intraclass Correlation Coefficient) istatistiğinden faydalanılmıştır.

Araştırmanın üçüncü alt amacı yargıcılardan Angoff (1-0) ve Nedelsky Yöntemleri ile elde edilen minimum geçme puanları (MGP) arasında uyum olup olmadığının incelenmesinde ise Pearson Momentler Çarpımı Korelasyon Katsayısı kullanılmıştır. Angoff (1-0) ve Nedelsky ile belirlenen kesme puanlarının ortalamaları arasında farka ise bağımlı gruplar için t testi ile bakılmıştır.

Araştırmanın dördüncü alt amacı, ROC ve aralık tahminine göre sınır değer saptama yöntemleri ile belirlenen kesme puanları arasında uyum olup olmadığının incelenmesi için öncelikle "Kolmogorov - Smirnov tek örneklem" testi yapılmış ve test puanlarının başarılı ve başarısız gruplardaki normallikleri incelenmiştir. Başarılı ve başarısız olarak sınıflandırılan öğrencilerin gerçek durumlarını belirlemek amacı ile Matematik ve Türkçe akademik başarı not ortalamaları ölçüt alınarak sınır değer belirleme yöntemlerinden ROC analizi yapılmıştır.

## Bulgular ve Yorumlar

### 1. Angoff (1-0) ve Nedelsky yöntemleri ile belirlenen kesme puanlarına göre başarılı/başarısız olarak kabul edilen öğrenci yüzdeleri arasında manidar bir fark var mıdır?

Araştırmanın birinci alt amacını cevaplandırmak için 40 yargıcıdan Angoff (1-0) ve Nedelsky Yöntemleri için belirlenen yönergelere göre 30 test maddesi hakkında karar vermeleri istenmiştir. Yargıcılara ait bu kararların ortalaması ile minimum geçme puanları hesaplanmıştır. Her iki yöntem için yargıcılara ait MGP'leri ve hesaplanan kesme puanları Tablo 1'de görülmektedir.

**Tablo 1.** Angoff (1-0) ve Nedelsky Yöntemi ile Belirlenen Türkçe ve Matematik Testi Kesme Puanları:

Yargıcı	Angoff (1-0) Türkçe Testi	Nedelsky Türkçe Testi	Angoff (1-0) Matematik Testi	Nedelsky Matematik Testi
Y1	12	5,73	9	8,9
Y2	11	6,40	12	7,73
Y3	10	6,48	11	7,98
Y4	9	6,82	10	7,81
Y5	8	7,07	10	7,73
Y6	8	6,98	13	8,23
Y7	10	7,15	11	8,15
Y8	7	7,65	11	7,31
Y9	11	8,15	12	9,15
Y10	11	8,82	10	7,14
Kesme Puan	9,70	7,12	10,90	8,01

Tablo 1'de görüldüğü gibi yargıcıların maddelere verdikleri kararların farklılığı yargıcılara ait minimum geçme puanlarının da farklı olmasına neden olmuştur. Türkçe ve Matematik testi için sırasıyla Angoff (1-0) Yöntemi ile belirlenen en yüksek kesme puanları (9, 70 ve 10, 90) iken, Nedelsky Yöntemi'yle daha düşük kesme puanları elde edilmiştir (7, 12 ve 8, 01). Angoff (1-0) ve Nedelsky Yöntemleri'ne göre elde edilen kesme puanlarındaki bu farklılık başarılı kabul edilen öğrenci sayılarını ve yüzdelerini de etkilemiştir. Yöntemlere ait kesme puanlarına göre

başarılı kabul edilen öğrenci sayıları ve yüzdeleri Tablo 2’de verilmiştir.

**Tablo 2.** Farklı Yöntemlerle Belirlenen Kesme Puanları, Başarılı Kabul Edilen Öğrenci Sayıları ve Yüzdeleri:

	Yöntem	Kesme Puanı	Başarılı Kabul Edilen Öğrenci Sayısı	Başarılı Kabul Edilen Öğrenci Yüzdesi
<b>Türkçe</b>	Angoff (1-0)	9,70	179	0,85
	Nedelsky	7,12	203	0,96
<b>Matematik</b>	Angoff (1-0)	10,90	109	0,52
	Nedelsky	8,01	169	0,80

Tablo 2’de görüldüğü gibi yöntemlere ait belirlenen kesme puanlarına göre araştırma grubunu oluşturan 211 öğrenciden başarılı kabul edilen öğrenci sayıları ve yüzdeleri farklılık göstermektedir. Türkçe ve Matematik testleri için en düşük kesme puanlarının belirlendiği Nedelsky Yöntemi, Angoff (1-0) Yöntemi’ne göre başarılı kabul edilen öğrenci sayısı ve yüzdesinin de en fazla olduğu yöntemdir.

Tablo 3 incelendiğinde yöntemlere ait kesme puanlarına göre başarılı öğrenci yüzdeleri ikili olarak karşılaştırıldığında tüm yüzdeler arasında 0,05 düzeyinde, manidar bir farklılık olduğu görülmektedir.

**Tablo 3.** Kesme Puanları, Başarılı Sayılan Öğrenci Sayıları, Yüzdeleri ve Z Değerleri

Testler	Karşılaştırılan Yöntemler	Kesme Puanı	Başarılı Sayılan Öğrenci Sayısı	Başarılı Sayılan Öğrenci Oranı	Z
<b>Türkçe</b>	Angoff (1-0)	9,70	179	0,85	4,91**
	Nedelsky	7,12	203	0,96	
<b>Matematik</b>	Angoff (1-0)	10,90	109	0,52	7,77**
	Nedelsky	8,01	169	0,80	

Başarılı olan öğrenci yüzdelerindeki bu manidar farklılıkların nedeni Angoff (1-0) ve Nedelsky yönteminde de verilen kararların farklı dayanakları olmasına bağlanabilir. Angoff (1-0) yönteminde yargıcılardan geçme/kalma sınırındaki öğrencinin maddeyi doğru yanıtlayıp yanıtlayamayacağını belirlemeleri istenirken; Nedelsky Yöntemi’nde geçme/kalma sınırındaki öğrencinin bu maddede eleyebileceği yanlış seçenek sayısını belirlemeleri istenmektedir. Bu kararlar doğrultusunda da minimum geçme puanları ortalaması alınarak kesme puanları hesaplanmaktadır. Böylece, yöntemlerin işleyiş süreçlerindeki farklılıklar kesme puanlarına yansımaktadır. Kesme puanlarındaki farklılık da başarılı öğrenci yüzdelerini etkilemektedir.

## 2. Bir güvenilirlik kanıtı olarak, kesme puanlarının belirlenmesinde;

- Angoff (1-0) yöntemi kullanıldığında yargıcılar arası uyum katsayısı nedir?
- Nedelsky yöntemi kullanıldığında yargıcılar arası uyum katsayısı nedir?

Araştırmanın ikinci alt amaç sorusunu cevaplandırmak için Türkçe ve Matematik testlerinde öncelikle Angoff (1-0) Yöntemi için karar vermiş olan yargıcıların, her bir test maddesi için verdikleri kararlarının uyumuna Cochran Q Testi ile bakılmıştır. Bu katsayı ile, bir kişiye ait gözlem değerlerinin diğer kişilerin gözlem değerlerine ne ölçüde benzer olduğu belirlenmiştir. Cochran Q Testi sonuçlarına göre Angoff (1-0) Yöntemi’nde Türkçe ve Matematik maddelerine ait yargıcı kararları arasında manidar bir uyum olmadığı belirlenmiştir ( $Q= 10,89, Q= 7,89, p > 0,05$ ). Angoff (1-0) Yöntemi’ne göre yargıcılar birbirlerinden farklı kararlar vermişlerdir, başka bir ifadeyle yargıcılar arası uyumun sağlanmadığı sonucuna ulaşılmıştır.

Yargıcıların kararlarındaki uyumsuzluklar, sınıf öğretmenliği programındaki öğrencilere

ilişkin beklentilerinin kendi aralarında farklılık göstermesine ve maddelerin güçlükleri ile ilgili kararlarının ve yargıcıların sahip oldukları alan bilgilerinin farklılıklarından da kaynaklanabilir. Güvenirlik kanıtı olarak, kesme puanlarının belirlenmesinde Nedelsky Yöntemi kullanıldığında yargıcılar arası uyum katsayısının ne olduğunu cevaplandırmak için yapılan Sınıf İçi Korelasyon Katsayısı Analizi kullanılmıştır. Nedelsky yöntemine göre kesme puan belirlemede incelenen sınıf içi korelasyon katsayıları Türkçe ve Matematik testleri için sırasıyla  $r=0,84$  ve  $r=0,98$ 'dir. Bu sonuçlar yargıcı kararları arası uyumun "mükemmel" olduğu şeklinde yorumlanabilir. Sınıf içi korelasyon katsayılarının büyük çıkması sonucunda, Nedelsky Yöntemi'nde yer alan yargıcı grubu içinde değişkenliğin olmadığı ancak grup ortalamaları arasında farkın olduğu sonucuna ulaşılmıştır.

### 3. Angoff (1-0) ve Nedelsky yöntemleri arasında, yargıcılardan elde edilen minimum geçme puanları (MGP) bakımından uyum var mıdır?

Araştırmanın üçüncü alt amaç sorusunu cevaplandırmak için öncelikle yöntemlere ait MGP'lerinin normal dağılıma sahip olup olmadığına bakılmıştır. Bu amaçla MGP'lerine ait betimsel istatistikler hesaplanmış, daha sonra normallik testi yapılmış ve sonuçlar Tablo 4'de özetlenmiştir.

Tablo 4'de görüldüğü gibi Angoff (1-0) ve Nedelsky yöntemlerinin Türkçe ve Matematik sorularına yönelik betimsel istatistikleri incelendiğinde Angoff (1-0) ve Nedelsky yöntemlerinin hem Türkçe, hem de Matematik minimum geçme puanlarına ait ortalama, ortanca ve tepe değerlerinin birbirine yakın olduğu görülmektedir. Bu bulgulara göre iki yöntem için de elde edilmiş olan kesme puanları dağılımının normal olduğu sonucuna varılmıştır. Kolmogorov-Smirnov testi, veri sayısının 50'den fazla olduğu durumlarda kullanılan bir normallik testidir. Bu sınamaya göre de yöntemlere ait minimum geçme puanlar normal dağılımdan manidar farklılıklar göstermemektedir.

**Tablo 4.** MGP'larına Ait Betimsel İstatistikler ve Normallik Testi Sonuçları

Betimsel İstatistikler	Angoff (1-0) Yöntemi Testler		Nedelsky Yöntemi Testler	
	Türkçe	Matematik	Türkçe	Matematik
N	10	10	10	10
Ortalama	9,70	10,90	7,12	8,01
Ortanca	10,00	11,00	7,02	7,89
Tepe Değer	11,00	10,00	5,73	7,73
Standart Sapma	1,63	1,19	0,89	0,63
Varyans	2,67	1,43	0,80	0,40
Çarpıklık Katsayısı	-0,35	0,23	0,52	0,61
Basıklık Katsayısı	-1,09	-0,36	0,32	-0,07
Kolmogorov-Smirnov İstatistiği	0,18	0,17	0,18	0,16
Kolmogorov-Smirnov Sig.	0,20	0,20	0,20	0,06

Minimum geçme puanları bakımından yöntemler arasındaki uyumun olup olmadığının incelenmesi için Pearson Momentler Çarpımı Korelasyon Katsayısı kullanılmıştır. Türkçe maddeleri için Angoff (1-0) ve Nedelsky Yöntemleri'ne ait minimum geçme puanları (MGP) arasında düşük düzeyde, manidar bir ilişkinin varlığı söz konusudur ( $r=0.30$ ,  $p<.05$ ). Başarılı sayılan öğrenci sayısı her iki yöntemde de belirlenen kesme puanlarına bağlı olarak birbirine yakın olmasına rağmen, bu katsayının düşük düzeyde olmasının nedeni, Angoff (1-0) Yöntemi'nde tüm sorular için evet ve hayır şeklinde kararlar verilirken; Nedelsky Yöntemi'nde maddelerin

elenebilir olma durumlarına göre karar verilmesi olabilir. Yargıcılar, Nedelsky yönteminde Türkçe maddeleri için kesme puanları belirlenirken elenebilir madde sayısını çoğu kez bir madde olarak belirlerken; Angoff (1-0) için ise yargıcılar elenebilirlikte sorun yaşanan maddelere 'evet' yanıtını daha kolay vermişlerdir. Bu durum aynı maddeler için kesme puanlarında farklılıklar oluşturmuştur. MGP'leri arasında Matematik maddeleri için Angoff (1-0) ve Nedelsky yöntemleri arasında orta düzeyde bir ilişki olduğu görülmektedir ( $r=0.63$ ,  $p<.05$ ). Buna göre, Matematik maddelerinde Türkçe'den daha yüksek kesme puanları belirleyen iki yöntem arasında MGP'ler bakımından yöntemler arasındaki ilişkinin daha yüksek olduğu görülmektedir. Pearson Momentler Çarpımı Korelasyon Katsayısı ortalamalardan bağımsız bir hesaplama yaptığı için yöntemlere ait MGP'lerin ortalamaları bağımlı gruplar T testi ile karşılaştırılmıştır ve sonuçları Tablo 5'de sunulmaktadır.

**Tablo 5.** Matematik ve Türkçe Test Puanlarının T Testi Sonuçları.

Karşılaştırılan Yöntemler		$\bar{X}$	t	sd	P
Matematik	Angoff (1-0) Nedelsky	10,90	7,12	9	0,00
		8,01			
Türkçe	Angoff (1-0) Nedelsky	9,70	4,15		0,00
		7,12			

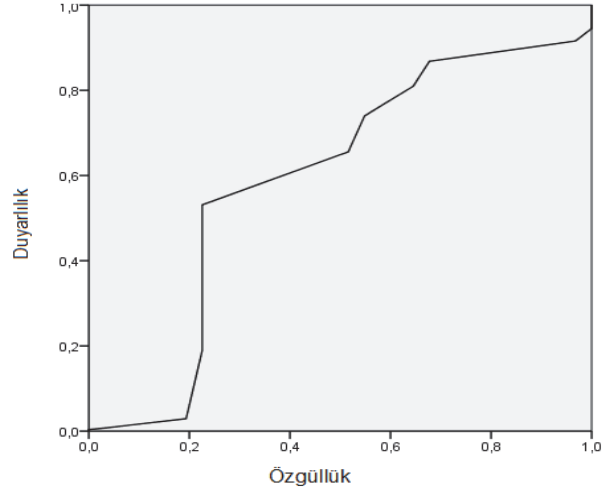
Tablo 5'de görüldüğü gibi Matematik soruları için Angoff (1-0) ve Nedelsky yöntemlerine ait minimum geçme puanları (MGP) ortalamaları sırayla 10,90 ve 8,01'dir. Yapılan t testi sonucuna göre bu iki ortalama arasındaki farkın manidar olduğu görülmektedir [ $t(9)=7,12$ ,  $p<.01$ ]. Türkçe soruları için Angoff (1-0) ve Nedelsky Yöntemleri'ne ait MGP'leri ortalamaları sıra ile 9,70 ve 7,12'dir. Bu iki ortalama arasındaki farkın da manidar olduğu görülmektedir [ $t(9)=4,15$ ,  $p<.01$ ].

#### 4. Sınır değer saptama yöntemleri olan ROC ve aralık tahmini ile belirlenen kesme puanlar nelerdir ve bu puanlar arasında uyum var mıdır?

Araştırmanın dördüncü alt amaç sorusunu cevaplandırmak için öncelikle araştırma grubunda yer alan öğrenciler Türkçe ve Matematik test puanlarına ve Türkçe, Matematik lisans akademik başarı not ortalamalarına göre başarılı ve başarısız olarak gruplandırılmış ve gruplarda dağılımın normal olup olmadığı Kolmogorov-Smirnov Tek Örneklem testi ile sınanmıştır. Araştırmada test sonucunda başarılı ve başarısız öğrencilerin bulunduğu grupların normal dağılım gösterdiği belirlenmiştir. Kolmogorov-Smirnov İstatistiği başarılı ve başarısız gruplarda (0, 14 ile 0, 27) arasında değerler almıştır ( $p> 0,01$ ). Betimsel istatistiklerin birbirine yakın değerlerde olması gruplarda normal dağılım varsayımının karşılandığını göstermektedir. Türkçe ve Matematik testi puanlarına göre başarılı ve başarısız olarak sınıflandırılan öğrencilerin gerçek durumlarını belirlemek amacıyla, önce Sınıf Öğretmenliği Programı birinci sınıf Matematik dersi akademik başarı not ortalamaları ölçüt alınarak sınır değerleri belirleme yöntemlerinden ROC analizi yapılmıştır. Bunun için, başarılı/başarısız öğrencileri doğru olarak ayıran kesim noktaları oluşturmak ve belirli bir kesim noktasını ölçüt almak gerekmektedir. ROC eğrisi yöntemi Duyarlılık, Özgüllük, Doğruluk oranları gibi oranları da kullanmaktadır. Araştırma grubunu oluşturan 211 öğrenci Matematik dersi not ortalamalarına göre (50 ile 59/ F-D) puan aralığı altında yer alma durumlarına göre başarılı ve başarısız olarak sıralanmış ve ROC eğrisi oluşturulmuştur.

### ROC Eğrisi

ROC analizi sonucunda, işlem karakteristik eğrisi altında kalan alan (AUC)=0.79 olarak hesaplanmış ve manidar ( $p<0.05$ ) bulunmuştur. Bu değer Matematik testinin öğrenci başarısını belirlemede ayırım yapabildiğini yani başarılı ve başarısız öğrencileri % 79 oranında (orta düzeyde) doğru sınıflandırabildiğini göstermektedir. Bu test için kesme puanı olarak hangi değer alınmalı sorusunun yanıtını vermek için, analiz sonucunda verilen her bir duyarlılık ve 1- özgüllük değerleri incelenmiş ve optimum nokta seçilmiştir. Bir başka deyişle duyarlılığın yüksek, 1-özgüllüğün düşük olduğu en iyi nokta belirlenmiştir. Şekil 1’de görüldüğü gibi, bu kesme değerinde eğrinin önce pik ve ardından plato yaptığı görülmektedir. Tablo 6’da ROC analizi ile belirlenen kesme puan için koordinatlar verilmiştir.



**Fig 1.** Matematik Testine İlişkin ROC Eğrisi Altında Kalan Alan

**Tablo 6.** Matematik Testi’ne İlişkin ROC Analizi ile Belirlenen Kesme Puan

ROC Eğrisi Koordinatları		
Kesme Puan	Duyarlılık	Özgüllük
2,00	1,000	1,000
3,50	,993	1,000
4,50	,945	1,000
5,50	,916	,968
6,50	,868	,677
7,50	,810	,645
8,50	,740	,548
9,50	,656	,516
10,50	,396	,226
11,50	,333	,226
12,50	,190	,226
13,50	,029	,194
14,50	,007	,032
15,50	,004	,000
16,50	,000	,000

Tablo 6’da kesme puanı olarak 8.50 seçilirse duyarlılık % 74 ve 1- özgüllük ise % 54,8 olarak elde edilir. Bu değer kesme puanı olarak belirlenmesinin sebebi, Matematik testi için duyarlılığın ve özgüllüğün her ikisinin de aynı anda en iyi değer aldığı puan olmasıdır. Kolmogorov-Smirnov istatistiğinin başarılı ve başarısız gruplarda (0, 14 ile 0, 27) arasında değerler alması ve normallik varsayımının sağlanması sonrasında kullanılan diğer sınır değer belirleme yöntemi olan aralık tahmini ile standart belirlemede “*Aritmetik Ortalama*  $\pm (1, 96) \times \text{Standart Sapma}$ ” formülü kullanılmıştır ve Matematik testinin üst ve alt sınırı belirlenmiştir. Böylece tahmini kesme puanlarının bulunduğu aralık saptanmıştır. Bu sınırlarda Matematik test puanlarına ait duyarlılık ve özgüllükler hesaplanmıştır. Shapiro (1999)’nun belirttiği gibi, duyarlılığı ve özgüllüğü yüksek olan test değeri, kesme puanı olarak belirlenmiştir. Matematik testinden başarılı olan n=191 öğrenci ve başarısız olan n=20 öğrencinin derse ilişkin akademik başarı not ortalamaları ölçüt alınarak belirlenen başarılı/başarısız olma durumlarına göre, başarılı gruptan belirlenen testin üst ve alt sınırlarının saptanmıştır.

Testin üst sınırı 16,19 ve alt sınırı ise 3,19 olarak belirlenmiştir. Bulunan bu kesim değerleri olan 16,19 ( $\approx 16$ ) ile 3,19 ( $\approx 3$ ) arasındaki kesim noktaları için duyarlılık ve özgüllükler hesaplanmıştır.

Tablo 7’de Matematik testi için tahmini kesim noktalarının duyarlılık ve özgüllükleri verilmiştir.

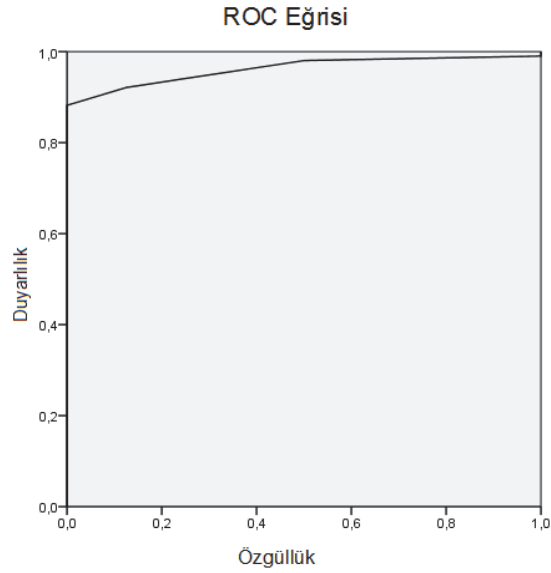
**Tablo 7.** Matematik Testi'ne Ait Tahmini Kesim Noktalarının Duyarlılık ve Özgüllükleri

Kesim (tahmini) değeri	Duyarlılık (%)	Özgüllük (%)
$\geq 3$	1,00	0,00
$\geq 8$	0,25	0,80
$\geq 13$	0,13	0,96

Matematik testi için tahmini kesme puanı için üst sınır 16 ve alt sınır ise 3 olarak belirlenmiştir. Bu nedenle sınırlar arasında belirlenen dört tahmini kesme puanına göre duyarlılık ve özgüllükler hesaplanmıştır. Tablo 7'de görüldüğü gibi 3,8 ve 13 kesme puanlarına göre hesaplanan duyarlılık ve özgüllükler incelendiğinde en yüksek duyarlılığın ve özgüllüğün sınır değer yaptığı kesme puanın 8-13 aralığında olduğu görülmüştür. Kesme puanı 16 değeri olarak belirlendiğinde duyarlılık çok küçülüp, özgüllük büyüdüğü için Tabloda yer verilmemiştir. Matematik testi için aralık değere göre belirlenen kesme puanı 8-13 aralığında yer almıştır. ROC Analizi'nde belirlenen kesme puan (8.50) ile bulunan bu aralık değer, sınır değer saptama yöntemlerine göre kesme puanı belirlemede uyumun olduğunu göstermektedir.

Türkçe testi için yapılan Kolmogorov-Smirnov istatistiği ile başarılı ve başarısız gruplarda normallik varsayımı sağlanmıştır. Türkçe testi için sınır değer saptama yöntemleri ROC analizi ve aralık tahmini ile kesme puanlarının belirlenmesi işlemleri sırasıyla aşağıda yer almaktadır. Sınıf öğretmenliği programı Türkçe dersi akademik başarı not ortalamaları ölçüt alınarak sınır değer belirleme yöntemlerinden İşlem Karakteristik Eğrisi (ROC) analizi yapılmıştır. Bunun için, başarılı/başarısız öğrencileri doğru olarak ayıran kesim noktaları oluşturmak ve belirli bir kesim noktasını ölçüt almak gerekmektedir. Uygun kesim noktasının belirlenmesi için akademik başarı not ortalamalarına göre (50 ile 59/ F-D) puan aralığı altında yer almış olan başarılı ve başarısız diye ayrılmış öğrencilerden elde edilen veriler kullanılmıştır.

ROC analizi sonucunda, işlem karakteristik eğrisi altında kalan alan  $AUC=0.96$  olarak hesaplanmış ve manidar bulunmuştur ( $p<0.05$ ). Bu değer Türkçe testinin öğrenci başarısını belirlemede mükemmel olduğunu, başarılı ve başarısız sayılan öğrencileri % 96 oranında doğru sınıflandırabildiğini göstermektedir. ROC eğrisinin altında kalan alan ne kadar büyük ise, bu testin doğru sınıflamada o kadar iyi olduğunu gösterir (Hanley, & McNeil, 1982; Grove, 2006). Bu test için kesme puanı olarak hangi değer alınmalı sorusunun yanıtını vermek için duyarlılığın yüksek, 1- özgüllüğün düşük olduğu en iyi nokta belirlenmiştir. Şekil 2'de görüldüğü gibi bu kesme değerinin önce eğrinin pik yaptığı, ardından plato yaptığı görülmektedir. Tablo 8'de ROC analizi ile belirlenen kesme puan koordinatları verilmiştir.

**Fig 2.** Türkçe Testi İlişkin ROC Eğrisi Altında Kalan Alan

**Tablo 8.** Türkçe Testi'ne İlişkin ROC Analiziyle Belirlenen Kesme Puanı.

ROC Eğrisi Koordinatları		
Kesme Puan	Duyarlılık	Özgüllük
3,00	1,000	1,000
5,50	,990	1,000
7,50	,980	,500
8,50	,921	,125
9,50	,882	,000
10,50	,759	,000
11,50	,542	,000
12,50	,350	,000
13,50	,108	,000
14,50	,030	,000
16,00	,000	,000

Tablo 8'de kesme puanı olarak 9,50 seçilirse duyarlılık % 96 ve 1-özgüllük ise % 0,00 olarak elde edilir. Bu değer kesme puanı olarak belirlenmesinin sebebi, Türkçe testi için duyarlılığın yüksek, 1- özgüllüğün düşük olduğu en iyi noktanın yer aldığı puan olmasındandır.

Sınır değer belirleme yöntemi olan aralık tahmini ile standart belirleme "*Aritmetik Ortalama + (1,96) x Standart Sapma*" formülü yardımıyla Türkçe testinin üst sınırı belirlenmiştir. Başarısız grubun test puanları ile de "*Aritmetik Ortalama -(1, 96)\* Standart Sapma*" ifadesi kullanılarak testin alt sınırı saptanmıştır. Bu sınırlarda Türkçe test puanlarına ait duyarlılık ve özgüllükler hesaplanmıştır. Türkçe testinden başarılı olan n=203 öğrenci ve başarısız olan n=8 öğrencinin derse ilişkin akademik başarı not ortalamaları

ölçüt alınarak belirlenen başarılı/başarısız olma durumlarına göre, başarılı gruptan belirlenen testin üst sınırı 15,24 ve alt sınırı ise 9,07 olarak belirlenmiştir. Bulunan bu kesim değerleri olan 15,24 ( $\approx 15$ ) ile 9,07 ( $\approx 9$ ) arasındaki kesim noktaları için duyarlılık ve özgüllükler hesaplanmıştır. Tablo 9'da Türkçe testi için tahmini kesim noktalarının duyarlılık ve özgüllükleri verilmiştir. Türkçe testi için üst sınırı 15 ve alt sınırı ise 9 olarak belirlenmiştir. Bu nedenle sınırlar arasında belirlenen üç tahmini kesme puanına göre duyarlılık ve özgüllükler hesaplanmıştır.

**Tablo 9.** Türkçe Testi'ne Ait Tahmini Kesim Noktalarının Duyarlılık ve Özgüllükleri

Kesim (tahmini) değeri	Duyarlılık (%)	Özgüllük (%)
$\geq 9$	0,77	0,48
$\geq 12$	0,46	0,80
$\geq 15$	0,14	0,98

Tablo 9'da görüldüğü gibi 9, 12, 15 kesme puanına göre hesaplanan duyarlılık ve özgüllükler incelendiğinde en yüksek duyarlılığın ve özgüllüğün sınır değer yaptığı kesme puanın 9-12 aralığında olduğu görülmüştür. Türkçe testi için aralık değere göre belirlenen kesme puanı 9-12 aralığında yer almıştır. ROC analizinde belirlenen kesme puan (9,50) ile bulunan bu aralık değeri, Sınır Değer Saptama Yöntemleri'ne göre kesme puanı belirlemede uyumun olduğunu göstermektedir. Sınır değer saptama yöntemlerinde ROC analizi ile belirlenen kesme puanlar incelendiğinde, Matematik testi için belirlenen kesme puan aralık tahmini ile belirlenen kesme puan sınırları içerisinde yer almıştır. Türkçe Testi için belirlenen kesme puan da aralık tahminine göre belirlenen sınır değer içerisinde yer almış ve uyum göstermiştir. ROC ve aralık tahminine göre yapılan kesme puanı belirleme yöntemleri arasındaki uyum birbirlerinin alternatifi olarak kullanılacaklarını da göstermiştir. ROC Analizi'nde kesme puanı belirlenirken, aralık tahmininde kesin değere ulaşmak için daha çok duyarlılık ve özgüllük hesaplamasının yapılması gerekmektedir ancak bu hesaplamalar yapılabilir de aralık tahmininde kesin bir puan belirlemek güçtür ve yorumlama açısından da fazla bilgi vermemektedir. ROC Eğrisi'nin kesme puanı belirleme kolaylığının yanı sıra eğri altında kalan alanın yorumlanabilmesi ile de grupları, olayları sınıflandırma; ayırt edebilme kolaylığı sağlamakta olduğu sonucuna ulaşılmıştır. Matematik

testinde İşlem Karakteristik Eğrisi altında kalan alan (AUC) başarılı/başarısız öğrencileri ayırmada orta düzeyde başarıyla; Türkçe testinde AUC mükemmel bir şekilde öğrencileri sınıflandırabilmiştir. ROC Eğrisi altında kalan alanın yorumlanmasıyla öğrenciler hakkında daha doğru sınıflandırma kararlarının da verilebileceği görülmüştür.

Angoff (1-0) ve Nedelsky Yöntemleri ile belirlenen kesme puanlarına göre başarılı/başarısız olarak kabul edilen öğrenci yüzdeleri arasında manidar bir fark olup olmadığına ilişkin bulgular incelendiğinde, Türkçe ve Matematik testleri için en düşük kesme puanlarının belirlendiği Nedelsky Yöntemi, Angoff (1-0) Yöntemi'ne göre başarılı kabul edilen öğrenci sayısı ve yüzdesinin en fazla olduğu yöntemdir. Angoff (1-0) ve Nedelsky yöntemleri için başarılı olan öğrenci yüzdeleri açısından elde edilen bu sonuç Tanrıverdi (2006)'nin araştırması ile örtüşmektedir. Araştırmada, yöntemler arası uyum incelendiğinde Türkçe dersi için Angoff ve Nedelsky yöntemleri ile Nedelsky ve Angoff (1-0) Yöntemleri arasında pozitif ve orta düzeyde uyum görülmüştür. Matematik dersi için, Angoff ve Angoff (1-0) Yöntemleri arasında pozitif ve orta düzeyde bir uyum olduğu saptanmıştır. Gündeğer (2012)'in araştırmasında da en yüksek kesme puanı Angoff (1-0) ile belirlenirken, en az başarılı kabul edilen öğrenci sayısı ve yüzdesi ise Angoff ve Ebel yöntemlerine göre en yüksek kesme puanı elde edilen Angoff (1-0) yöntemi olmuştur. Bu durum Impara ve Plake (1997)'in belirttiği gibi yargıcıların Angoff (1-0) Yöntemi'nde daha kolay karar vermesi ve Nedelsky Yöntemi'nde ise karar vermede elenebilecek seçenekleri belirlemenin daha zor olduğu şeklindeki bulguları ile de örtüşmektedir. Bu bulguların yanı sıra, araştırmada Türkçe ve Matematik derslerinin sayısal ve sözel içerikli olması ile geçme puanı üzerinde puan alan öğrenci yüzdeleri ikili olarak karşılaştırıldığında bu yüzdeler arasında manidar bir ilişki olduğu görülmektedir. Bu durum ise derslerin sayısal ya da sözel içerikli olması ile kesme puanı üzerinde puan alan öğrenci yüzdeleri arasında farklılıklar olduğu şeklinde ifade edilebilir.

Bir güvenilirlik kanıtı olarak, kesme puanlarının belirlenmesinde; Angoff (1-0) ve Nedelsky yöntemi kullanıldığında yargıcılar arası uyum katsayısının ne olduğuna ilişkin bulgular incelendiğinde, Cochran Q testi sonuçlarına göre Angoff (1-0) Yöntemi'nde Türkçe ve Matematik maddelerine ait yargıcı kararları arasında manidar bir uyum olmadığı, Nedelsky Yöntemi'ne göre kesme puan belirlemede incelenen sınıf içi korelasyon katsayılarına göre Türkçe ve Matematik testlerinde yargıcı kararları arası uyumun "mükemmel" olduğu görülmektedir. Bu sonuç Ömür ve Selvi (2010)'nin Angoff, Ebel ve Nedelsky Yöntemleri ile ve farklı yargıcı sayıları kullanılarak elde edilen kesme puanlarının sınıflama tutarlılığı indekslerinin incelendiği araştırması ile örtüşmektedir. Yargıcı puanlamaları sonucunda, üç yöntemde ve üç ayrı yargıcı sayısında en yüksek sınıflama tutarlılığı indeksinin Nedelsky yöntemi ile elde edildiği görülmüştür. Chang (1999) yaptığı araştırmada Nedelsky ve Angoff Standart Belirleme Yöntemleri'ni karşılaştırmış; Ömür ve Selvi'nin araştırma sonuçlarına benzer şekilde Nedelsky Yönteminde Angoff Yöntemine göre yargıcılar arası uyumsuzluğun daha düşük düzeyde olduğunu belirlemiştir. Angoff yöntemi yargıcılar için puanlama kolaylığı sağlamakla birlikte Nedelsky Yöntemi'nde olduğu gibi öğrencinin hangi seçeneği ya da seçenekleri eleyebilir ya da eleyemez şeklindeki kararları verilmemesi neden ile yargıcı kararları bakımından uyumun düşük olduğu yöntem olduğu ifade edilebilir.

Angoff (1-0) ve Nedelsky Yöntemleri arasında, yargıcılardan elde edilen minimum geçme puanları (MGP) bakımından uyum olup olmadığına ilişkin bulgular incelendiğinde, Türkçe maddeleri için Angoff (1-0) ve Nedelsky Yöntemleri'ne ait minimum geçme puanları (MGP) arasında düşük düzeyde, manidar bir ilişkinin varlığı söz konusudur. MGP'leri arasında Matematik maddeleri için Angoff (1-0) ve Nedelsky Yöntemleri arasında orta düzeyde bir ilişki olduğu görülmektedir. Türkçe ve Matematik maddeleri için Angoff (1-0) ve Nedelsky Yöntemleri'ne ait minimum geçme puanları (MGP) ortalamaları arasındaki farkın manidar olduğu görülmektedir. Çetin ve



Gelbal (2010) arařtırmalarında Türkçe için Nedelsky ve Angoff (1-0) Yöntemleri arasında manidar bir fark görüldüğü; Matematik için Nedelsky ve Angoff (1-0) Yöntemleri arasında ise manidar bir fark görülmediği sonucuna ulaşmışlardır. Türkçe için Nedelsky ve Angoff (1-0) Yöntemleri arasında ise pozitif ve orta düzeyde ilişki görülmüş iken; Matematik için Nedelsky ve Angoff (1-0) Yöntemleri arasında pozitif ve düşük düzeyde manidar bir ilişki görülmüştür. Bir diğer arařtırmada ise Gündeğer (2012), Angoff ve Angoff (1-0) yöntemleri ile belirlenen minimum geçme puanları (MGP) arasında pozitif ve orta düzeyde bir ilişki olduğunu ve iki yönteme ait minimum geçme puan ortalamaları arasında manidar bir farklılığın olmadığını belirlemiştir. Bu arařtırma bulguları yöntemler arasında ilişkinin olmasının yanı sıra kullanılan test maddelerine ve yargıcı özelliklerine baėlı olarak minimum geçme puan ortalamaları bakımından farklılıkların olduğunu da göstermektedir.

Sınır deėer saptama yöntemleri olan ROC ve aralık tahmini ile belirlenen kesme puanlarının neler olduėu ve bu puanlar arasında uyumun olup olmadığına ilişkin bulgular incelendiğinde, ROC analizi ile Türkçe ve Matematik testi için belirlenen kesme puanlar aralık tahmini ile belirlenen kesme puan sınırları içerisinde yer almıştır. ROC ve aralık tahminine göre yapılan kesme puanı belirleme yöntemleri arasındaki uyum birbirlerinin alternatifi olarak kullanılacaklarını da göstermiştir. Bu sonuç Çamlıca ve Dişçi (2008) tarafından yapılan arařtırma ile örtüşmektedir. ROC, aralık tahmini ve yüzde deėerlerine göre sınır deėer saptama yöntemlerini kullanarak hasta ve saėlıklı kişilerin sınıflandırma doğrulukları karşılaştırıldığı arařtırmada bu üç yöntemin benzer sınıflandırmalar yaptığı da görülmektedir. Centor ve Schwartz (1985) arařtırmalarında ROC Eğrisi altında kalan alanı hesaplamış ROC eğrileri tutarlı sonuçlar vermiştir. Bu ve benzeri arařtırmalarda ROC analizinin ve eğrisi altındaki alanların aralık tahmini ve yüzde deėerlerine göre daha çok kullanılmasının sebebi belirli bir kesme puan verebilmesinden ve sınıflandırma ile ilgili daha doėru sonuçlar verebilmesi olarak yorumlanabilir.

### **Sonuç ve Öneriler**

Bu bölümde arařtırmanın sonuçlarına, uygulamaya yönelik öneriler ile ileride yapılacak arařtırmalara yönelik önerilere yer verilmiştir.

### **Sonuçlar**

- 1) Angoff (1-0), Nedelsky Yöntemleri ile belirlenen kesme puanlarının birbirinden farklı olduėu görülmektedir. Türkçe ve Matematik testi için Angoff (1-0) yöntemiyle elde edilirken,
- 2) Nedelsky Yöntemi'yle daha düşük kesme puanları belirlenmiştir. Kesme puanlarındaki bu farklılık başarılı kabul edilen öğrenci sayılarını ve yüzdelerini de deėiştirmiştir. Yöntemlere ait kesme puanlarına göre başarılı kabul edilen öğrenci yüzdeleri arasında da manidar farklılık olduėu sonucuna ulaşılmıştır.
- 3) Güvenirlilik kanıtı olarak, kesme puanlarının belirlenmesinde; Angoff (1-0) ve Nedelsky Yöntemleri'nde yargıcılar arası uyum arařtırılmıştır. Angoff (1-0) yönteminde Türkçe maddelerine ait yargıcı kararları arasında uyum olmadığı yine Angoff (1-0) Yöntemi'nde Matematik maddelerine ait yargıcı kararları arasında da verilen kararların uyumlu olmadığı görülmüştür. Bu durum; yargıcıların sınıf öğretmenliği programındaki öğrencilere ilişkin beklentilerinin kendi aralarında farklılık göstermesine ve maddelerin güçlüklerine ilişkin kararlarında yargıcıların sahip oldukları alan bilgilerinin farklılıklarına bağlanmıştır. Nedelsky Yöntemi'ne göre kesme puan belirlemede yargıcı kararları arası uyumun olduėu görülmüştür. Bu durumda, Nedelsky Yöntemi'nde yer alan yargıcı grubu içinde deėişkenliğin olmadığı, grup ortalamaları arasında farkın olduėunu sonucuna ulaşılmıştır.
- 4) Elde edilen minimum geçme puanları (MGP) bakımından Türkçe maddeleri için Angoff (1-0)

ve Nedelsky Yöntemleri arasında düşük düzeyde uyumun olduğu ve MGP ortalamalarının manidar düzeyde farklılaştığı sonucuna ulaşılmıştır. Matematik maddeleri için daha yüksek kesme puanları belirleyen Angoff (1-0) ve Nedelsky yöntemleri arasında minimum geçme puanları bakımından varolan ilişkinin ve uyumun daha yüksek olduğu görülmüştür.

- 5) ROC ve aralık tahminine göre Sınır Değer Saptama Yöntemleri ile belirlenen kesme puanları arasındaki uyum araştırılmıştır. Türkçe ve Matematik testleri için Sınır Değer Saptamaya Yönelik Kesme Puanlarının Belirlenmesi için yapılan ROC analizi sonucunda, Matematik testinin öğrenci başarısını belirlemede orta düzeyde olduğunu ve başarılı/başarısız olan öğrencileri doğru sınıflandırabildiğini göstermiştir. Bu değer, Türkçe Testinde öğrenci başarısını belirlemede mükemmel düzeyde olduğu belirlenmiştir. Bu durum; Türkçe testinde başarılı ve başarısız öğrencilerin doğru sınıflandırabildiğini göstermiştir. Matematik ve Türkçe Testi için aralık tahminine göre belirlenen kesme puanlarının ROC analizinde belirlenen kesme puanları ile uyumlu olduğunu göstermiştir.

Sonuç olarak, bu çalışmada Türkçe ve Matematik testleri için Angoff (1-0) ve Nedelsky; ROC ve Aralık Tahmini Yöntemleri'ne göre farklı kesme puanları belirlenmiştir. Hem Yargıcı Kararları ile hem Sınır Değerlerle Kesme Puanı Belirlemede Yöntemleri'nin birbirleri ile uyumlu olduğu sonucuna ulaşılmıştır. Kesme puanı belirleme sürecinde diğer yöntemlere göre puanlanması kolay olan Angoff (1-0) Yöntemi Nedelsky'e göre daha yüksek kesme puanları üretmiştir. Ancak yargıcılar arası uyum sağlanamamıştır. Nedelsky Yöntemi'nde yargıcılar arası uyumun daha yüksek olduğu görülmüştür. ROC Analizinin, çalışmalarda kesme puanı belirleme ve öğrencileri sınıflamada doğru kararlar verilmesinde yaygın olarak kullanılan kesme puanı belirleme yöntemleri olan Angoff (1-0) ve Nedelsky gibi kullanılabilen sonuçuna ulaşılmıştır.

## Öneriler

### 1. Araştırma Sonuçlarından Çıkan Öneriler

- Bu çalışmada Angoff (1-0) ve Nedelsky; İşlem Karakteristik Eğrisi ve Aralık Tahminine Göre Kesme Puanı Belirleme Yöntemleri kullanılmıştır. Benzer bir çalışma Ebel, Jeager gibi kesme puanı belirleme yöntemleri ile de yapılabilir.
- Bağlı değerlendirme yerine mutlak değerlendirmeyi esas alan testlerde kesme puanları arasında yer alan düzeylerin tanımlanması, bu kesme puanları açısından nicel bilginin yanı sıra hem öğrenciye hem sınavı yapan kurum/lara nitel bilgi de sağlayacağından dolayı, kesme puanları arasında kalan düzeylerin YGS, LYS gibi benzer sınavlarda belirlenmesi önerilebilir.
- Yargıcı kararlarına başvurmadan önce araştırmanın yapılacağı konu alanı ile ilgili bilgiye sahip büyük örneklem üzerinde çalışılması önerilebilir. Çalışmalarda kullanılacak Kesme Puan Belirleme Yöntemi ile doğru puanlama yapılması için ön deneme çalışmalarının yapılması da önerilebilir.
- Sınır değer saptama yöntemlerinden ROC analizinin iyi bir sınıflandırma analizi olarak kullanılabilen araştırma bulgularında ortaya çıkmıştır. Bu analizle öğrencilerin başarı, tutum, kaygı vb. pek çok özellikleri işlem karakteristik eğrisi altındaki alan kullanılarak yorumlanabilir.

### 2. Araştırmacılara Yönelik Öneriler

- Bu çalışmada standart belirleme sürecinde yargıcı sayısı iki test için toplam 40 yargıcıdan oluşmuştur. Bundan sonra yapılacak olan çalışmalarda Genellebilirlik Kuramı ile yargıcı sayısının belirlenmesine ilişkin bir çalışma yapılabilir. Genellebilirlik Kuramı değişkenler arasındaki değişik varyasyonları gerçekleştirebildiği için böyle bir çalışma bu konudaki bir eksikliği giderebilir.
- Bu çalışma öğrencilere uygulamada, yargıcılara puanlamada zorluk yaşatmaması için 30

maddelik bir test üzerinden yürütülmüştür. Yapılacak araştırmalarda madde sayısı daha fazla olan testlerle çalışılabilir.

- Bu araştırmada daha önce eğitim araştırmalarında kullanılmamış olan sınır değer saptama yöntemleri incelenmiştir. Bundan sonraki araştırmalarda eğitimin farklı konularında çalışılabilir. Özellikle ayırt etme, sınıflandırma amaçlı çalışmalarda iyi sonuçlar veren bu yöntemlerin ROC, sınıflandırma, kümeleme amaçlı kullanılan çok değişkenli istatistiksel tekniklerle birlikte kullanımı ile yeni araştırmaların yapılması önerilebilir.

- Alanyazında ölçek geliştirme ve uyarlamada geçerlik kanıtı elde etmek amacıyla sık olmasa da kullanılan ve iyi sonuçlar veren ROC ve bu eğri altındaki alan hesaplamalarından yararlanarak, eğitimde ölçme aracı geliştirme ve uyarlama çalışmalarında da yeni araştırmalar yapılabilir.

- Bu araştırmada sadece test puanı ile sınırlı olan başarı değerlendirilmesinin yerine öğrenci akademik başarı not ortalamalarının da ölçüt olarak alınmıştır. Bundan sonraki araştırmalarda başarı puanlarının ölçüt olarak alındığı, mutlak değerlendirmelerin yapıldığı yeni çalışmalar planlanabilir.

- Bu araştırmada, öğrenci akademik başarı not ortalamasının ve öğrencinin testten aldığı puanların birlikte değerlendirildiği ve böylece kesme puanının belirlendiği ROC analizi seçme, yerleştirme ve sınıflama amaçlı testlerin kullanıldığı araştırmalarda da kullanılabilir.

- Araştırmalarda bağlı değerlendirme yapılan sınavlar için Test Merkezli Yöntemlerle belirlenen kesme puanlarını kullanarak, hangi kesme puanının kullanımının ölçüt belirlemede daha uygun olduğu sorusunun yanıtı aranabilir.

#### Yazarın Notu

Bu araştırma, Doç. Dr. Çokluk danışmanlığında doktora tezi olarak Ankara Üniversitesi Eğitim Bilimleri Fakültesi, Ölçme ve Değerlendirme Bölümü'nde sunulmuştur (24.06.2013).

#### KAYNAKÇA

- Anastasi, A. (1988). *Psychological Testing*. (sixth edition). NY: Mc Millan Publishing Company.
- Büyüköztürk, Ş., Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2010). *Bilimsel Araştırma Yöntemleri*. 7. Baskı. Ankara: Pegem A Yayıncılık.
- Centor, R. M., & Schwartz, J. S. (1985). "An Evaluation of Methods for Estimating the Area under the Receiver Operating Characteristic (ROC) Curve". *Medical Decision Making Journal*, 5 (2), 149-156.
- Chang, L. (1999). "Judgmental Item Analysis of the Nedelsky and Angoff Standard-Setting Methods". *Applied Measurement in Education*, 12 (2), 151-65.
- Cizek, G. J. (2001). Conjectures on the Rise and Call of Standard Setting: An Introduction to Context and Practice. Ed. G. Cizek. *In setting Performance Standards: Concepts, Methods and Perspectives*, 3-17. Mahwah, N. J.: Erlbaum.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Thousand Oaks, CA: Sage Publications Ltd.
- Conway, M. J. (2006). *How to Collect Data: Measurement and Evaluation*. New York: American Society for Training and Development. Workplace Learning and Performance Press.
- Crocker, L., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. NY: CBS College Publishing Company.
- Cronbach, L. J. (1990). *Essentials of Psychological Testing*. (fifth edition). New York: Harper Collins Publishers Inc.

- Çamlıca, H., & Dişçi, R. (2008). "Tanı Testlerinde Sınır Değerlerinin Belirlenmesi". *Türk Onkoloji Dergisi*, 23 (1), 26-33.
- Çetin, S., & Gelbal, S. (2010). "Farklı Standart Belirleme Yöntemlerinin Geçme Puanları Üzerine Etkisi". *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 43 (1), 79-95.
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). "Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach". *Biometrics*, 44 (3), 837-845.
- Dirican, A. (1991). *ROC Eğrisi Çözümlemesi ile Tanı Testlerinin Değerlendirilmesi ve Bilgisayar Uygulaması*. Yayınlanmamış Doktora Tezi. İstanbul Üniversitesi Sağlık Bilimleri Enstitüsü. İstanbul.
- Goodwin, L. D. (1996). "Focus on quantitative methods determining cut-off scores". *Research in Nursing and Health*, 19 (3), 249-256.
- Grove, W. M. (2006). *Mathematical Aspects of Diagnosis*. United States of America, 50-75.
- Gündeğer, C. (2012). *Angoff, Yes/No ve Ebel Standart Belirleme Yöntemlerinin Karşılaştırılması*. Yayınlanmış Yüksek Lisans Tezi. Hacettepe Üniversitesi. Ankara.
- Hanley, J. A., & McNeil, B. J. (1982). "The Meaning and Use of The Area Under a Receiver Operating Characteristic (ROC) Curve". *Radiology*, 143, 29- 36.
- Hess, B., Subhiyah, R. G., & Giordano, C. (2007). "Convergence Between Cluster Analysis and the Angoff Method for Setting Minimum Passing Scores on Credentialing Examinations". *Evaluation and the Health Professions*, 30 (4), 362-375.
- Hurtz, G. M., & Hertz, N. R. (1999). "How Many Raters Should be Used for Establishing Cut off Scores with the Angoff Method? A Generalizability Theory Study". *Educational and Psychological Measurement*, 59 (6), 885-897.
- Impara, J. C., & Plake. B. S. (1997). "Standard setting: An alternative approach". *Journal of Educational Measurement*, 34 (4), 353-366.
- Irwin, P. (2007). *An Alternative Examinee - Centered Standard Setting Strategy*. Yayınlanmamış Doktora Tezi. University of Nebraska, USA.
- Jeager, R. M. (1989). Certification of Student Competence: Ed. R. L. Linn. *Educational Measurement*, 485-514. New York: Macmillan.
- Karasar, N. (2005). *Bilimsel Araştırma Yöntemi*. Ankara: Nobel Yayın Dağıtım.
- Kanık, E. A., & Erden, S. (2003). "Tanı Testlerinin Değerlendirilmesinde ROC Eğrisinin Kullanımı". *Mersin Üniversitesi Tıp Fakültesi Dergisi*, 3, 260-264.
- Karayianni, T., Tretiak, O. J., & Herrmann, N. (1996). *ROC analysis: Comparison between the Binormal and the Neyman-Pearson Model. Signals, Systems and Computers*, Conference Record of the Thirtieth Asilomar Conference. 3-6 Nov.
- Knapp, RG., & Miller III MC. (1992). "Clinical Epidemiology and Biostatistics". United States of America, Williams and Wilkins Press, Le CT. *Introductory Biostatistics*. United States of America, Wiley and Sons Pres, 336-337.
- Kubiszyn, T., & Borich, G. (2003). *Educational Testing and Measurement: Classroom Application and Practice*. USA: John Wiley and Sons, Inc.
- Letner, C. (1992). *Introduction to Statistics. Statistical Tables*. Geigy Scientific Tables, Ciba- Geigy Ltd. Basel-Switzerland.
- Livingston, S. A., & Zieky M. J. (1989). "A comparative study of standard- setting methods". *Applied Measurement in Education*, 2 (2), 121-141.
- Mathison, S. (2005). *Encyclopedia of Evaluation*. CA: Sage Publications, Inc.
- Metz, C. E., Herman, B. A., & Shen, J. H. (1998). "Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously- distributed data". *Statistics Medicine*, 17 (9), 1033-1053.
- Metz, C. E., & Pan, X. (1999). "Proper binormal ROC curves: Theory and maximum-likelihood estimation". *Journal of Mathematical Psychology*, 43 (1), 1-33.
- McClish, D. K. (1989). "Analyzing a portion of the ROC curve". *Medicine Decision Making Journal*, 9 (3), 190-5.

- Nedelsky, L. (1954). "Absolute Grading Standards for Objective Tests". *Educational and Psychological Measurement*, 14, 3-19.
- Obuchowski, N. A., Lieber, M. L., & Wians, F. H. (2004). "ROC Curves in Clinical Chemistry: Uses, Misuses and Possible Solutions". *Clinical Chemistry*, 50 (7), 1118-25.
- Ömür, S., & Selvi, H. (2010). "Angoff, Ebel ve Nedelsky Yöntemleriyle Belirlenen Kesme Puanlarının Sınıflama Tutarlılıklarının Karşılaştırılması". *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1 (2), 109-113.
- Özçelik, A. D. (1992). *Ölçme ve Değerlendirme*. Ankara: ÖSYM Yayınları.
- Pawar, M. S. (2004). *Data Collecting Methods and Experiences: A Guide for Social Researchers*. India: Sterling Publishers.
- Schulz, M. E. (2006). "Commentary: a response to reckase's conceptual framework and examples for evaluating standard setting methods". *Educational Measurement: Issues and Practice*, 25 (3), 4-13.
- Shapiro, D. E. (1999). "The interpretation of diagnostic tests". *Statistical Methods in Medical Research*, 8 (2), 113-34.
- Shen, L. (2001). "A comparison of Angoff and Rasch Model Based Item Map Methods in Standard Setting". *Paper Presented at the Annual Meeting of the American Educational Research Association*. Seattle, WA.
- Solberg, H. E. (1983). "The Theory of Reference Values. Statistical Treatment of Collected Reference Values. Determination of reference limits". *Journal of Clinical Chemistry and Clinical Biochemistry*, 21 (11), 749-60.
- Taga, Y., Aslan, D., & Güner, G. (2000). *Tıbbi Laboratuvarlarda Standardizasyon ve Kalite Yönetimi*. Ankara: Türk Biyokimya Derneği Yayınları.
- Tanrıverdi, S. (2006). *Standart Belirleme Yöntemlerinin Geçme Puanları Üzerine Etkisi*. Yayınlanmış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Ankara.
- Taşdelen, G. (2009). *Nedelsky ve Angoff Standart Belirleme Yöntemlerinin Genellenebilirlik Kuramı İle Karşılaştırılmasına İlişkin Bir Araştırma*. Yayınlanmış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Ankara.
- Turgut, F. M. (1992). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara: Saydam Matbaacılık.
- Zweig, M. H., & Campbell, G. (1993). "Receiver - operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine". *Clinical Chemistry*, 39, 561-577.
- Zwick, R., Şentürk, D., & Wang, J. (2001). "An investigation of alternative methods for item mapping in the national assessment of educational progress". *Educational Measurement: Issues and Practice*, 20 (2), 15-25.
- Wagner, R. F., Metz, C. E., & Campbell, G. (2007). "Assessment of Medical Imaging Systems and Computer Aids: A Tutorial Review". *Academic Radiology*, 14, 723-748.