



Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü

Eğitim Bilimleri Anabilim Dalı

Eğitimde Ölçme ve Değerlendirme Bilim Dalı

**NORM VE ÖLÇÜT DAYANAKLI DEĞERLENDİRMELERİN
KARŞILAŞTIRILMASINA İLİŞKİN BİR ÇALIŞMA**

İsmail ÇUKADAR

Yüksek Lisans Tezi

Ankara, 2013

NORM VE ÖLÇÜT DAYANAKLI DEĞERLENDİRMELERİN KARŞILAŞTIRILMASINA
İLİŞKİN BİR ÇALIŞMA

İsmail ÇUKADAR

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü

Eğitim Bilimleri Anabilim Dalı

Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Yüksek Lisans Tezi

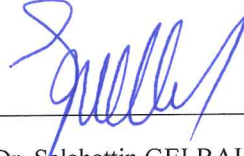
Ankara, 2013

KABUL VE ONAY

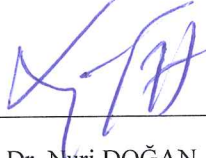
İsmail Çukadar tarafından hazırlanan “Norm ve Ölçüt Dayanaklı Değerlendirmelerin Karşılaştırılmasına İlişkin Bir Çalışma” başlıklı bu çalışma, 16.07.2013 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Yüksek Lisans Tezi olarak kabul edilmiştir.



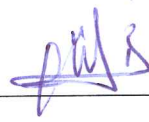
Doç. Dr. Mehtap ÇAKAN (Başkan)



Prof. Dr. Selahattin GELBAL (Danışman)



Doç. Dr. Nuri DOĞAN



Yrd. Doç. Dr. Burcu ATAR



Öğr. Gör. Dr. Sevda TANRIVERDİ ÇETİN

Yukarıdaki imzaların adı geçen öğretim üyelerine ait olduğunu onaylarım.

Prof. Dr. Yusuf ÇELİK

Enstitü Müdürü

BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kağıt ve elektronik kopyalarının Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezimin/Raporumun tamamı her yerden erişime açılabilir.
- Tezim/Raporum sadece Hacettepe Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin/Raporumun 1 yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

16.07.2013



İsmail ÇUKADAR

TEŞEKKÜR

Yüksek lisansım boyunca yolculuk etmenin ve işten izin almanın zor olduğu günlerde destekleriyle ve zevkli ders anlatımlarıyla bu sorunları bir kenara bırakmamı sağlayan, hem ders hem de tez aşamasında desteklerini hiç esirgemeyen değerli öğretmenim ve tez danışmanım Prof. Dr. Selahattin GELBAL'a sonsuz teşekkürlerimi sunarım.

Tez konusunun oluşmasında ve tezin şekillenmesinde en büyük paya sahip olan değerli öğretmenim Doç. Dr. Hülya KELEÇİOĞLU'na fikirleri ve destekleri için çok teşekkür ederim. Özellikle tezin yöntem kısmında büyük destek veren ve tezi diğer çalışmalarla karşılaştırarak tezin özgün olmasını sağlayan değerli öğretmenim Doç. Dr. Nuri DOĞAN'a çok teşekkür ederim.

Yüksek lisansım boyunca tüm derslerimin keyifli geçmesini sağlayan değerli arkadaşlarım Emre TOPRAK ve Ömür Kaya KALKAN'a çok teşekkür ederim. Üniversiteden uzak olmam nedeniyle okulla ilgili gelişmeler konusunda destek olan Ömür Kaya KALKAN'a ve tezin veri toplama sürecinde yardımcı olan Emre TOPRAK'a ayrıca tekrar teşekkür ederim.

Tezle ilgili gelişmeler konusunda bana destek olan, fikirlerini paylaşan ve yapılması gereken çalışmalarla ilgili yön gösteren değerli arkadaşım Sakine GÖÇER ŞAHİN'e çok teşekkür ederim.

Üniversiteye kayıt döneminde ve tez savunma sürecinde verdiği destekler nedeniyle arkadaşım Süleyman Özcan'a çok teşekkür ederim.

Yüksek lisans yapabilmem için gerekli izinleri almam konusunda bana destek olan ve yol gösteren eski okul müdürüm Sami CAN'a çok teşekkür ederim.

Beni Hacettepe Üniversitesi'ne yönlendiren ve nitelikli bir eğitim almamı sağlayan değerli öğretmenim Prof. Dr. Güzver YILDIRAN'a ve lisans hayatım boyunca beni ölçme ve değerlendirme dersleri almam için yönlendirerek bu alanı sevmemi sağlayan sevgili öğretmenim Yrd. Doç. Dr. Hande SART'a sonsuz teşekkürler.

Hep yanımda olarak bana destek olan, ben derslerle ilgilenirken evin tüm sorumluluklarını üstlenen, sevgisiyle kara bulutlarımı dağıtan ve tezin bu aşamaya gelmesinde en büyük paya sahip olan canım eşime ne kadar teşekkür etsem azdır...

ÖZET

ÇUKADAR, İsmail. *Norm ve Ölçüt Dayanaklı Değerlendirmelerin Karşılaştırılmasına İlişkin Bir Çalışma*, Yüksek Lisans Tezi, Ankara, 2013.

Eğitim sistemi çerçevesinde kurumlar, bireylerin bir kapsam alanına ya da yapıya ilişkin yeterliliklerini ve seviyelerini belirlerken farklı değerlendirme yöntemleri kullanmaktadır. Bu çalışmada norm ve ölçüt dayanaklı değerlendirmeler sonucunda alınan kararların kullanılan değerlendirme yöntemine göre değişip değişmediği, ilişkisel türden araştırma modeli ile incelenmiştir.

Veri toplama sürecinin ilk aşamasında, lisans düzeyinde ölçme ve değerlendirme dersine ilişkin geliştirilen 20 soruluk çoktan seçmeli başarı testi, 107 rehberlik ve psikolojik danışmanlık 2. sınıf öğrencisine uygulanmıştır. Uygulama sonucunda elde edilen ham puanlar, T puanına dönüştürülmüştür. T puanları, 50 ve 60 ölçütleri dikkate alınarak başarılı ve başarısız olacak şekilde sınıflandırılmıştır. İkinci aşamada ise ölçüt dayanaklı değerlendirme olan Angoff ve Nedelsky standart belirleme yöntemlerine ilişkin görüş almak için ölçme ve değerlendirme alanında uzman olan 14 kişiden faydalanılmıştır. Uzman görüşleri ile Angoff ve Nedelsky Yöntemleri doğrultusunda başarı ölçütünü gösteren iki farklı kesme puanı hesaplanmıştır. Başarı testinden elde edilen ham puanlar ise kesme puanları dikkate alınarak başarılı ve başarısız olacak şekilde sınıflandırılmıştır. Veri toplama sürecinin son aşamasında ise çalışmada uygulanan yöntemlerin uygunluk geçerliğini incelemek için öğrencilerin ölçme ve değerlendirme dönem sonu notları elde edilmiştir.

Araştırmada kullanılan yöntemlere ilişkin kararların farklılığını ortaya koymak için bağımlı iki oran arasındaki fark, z testi ile sınanmıştır. Kullanılan yöntemlerin ders notları ile uyumu ise Cohen'in Kappa testi ile incelenmiştir.

Başarılı öğrenci yüzdeleri, kullanılan değerlendirme yöntemine göre ikili olarak karşılaştırıldığında tüm değerlendirme yöntemleri arasında 0,01 düzeyinde anlamlı farklılıklar bulunmuştur. Bu doğrultuda norm ve ölçüt dayanaklı değerlendirmeler sonucu alınan kararlarda anlamlı farklılıklara ulaşılmıştır. Uygunluk geçerliği açısından ise sadece ölçüt dayanaklı değerlendirme olan Nedelsky Yöntemi'nin ders notlarıyla uyumlu olduğu, diğer yöntemlerin ise düşük uyuma sahip olduğu sonucuna ulaşılmıştır.

Anahtar Sözcükler

Norm Dayanaklı Değerlendirme, Ölçüt Dayanaklı Değerlendirme, Angoff Yöntemi, Nedelsky Yöntemi

ABSTRACT

ÇUKADAR, İsmail. *A Study Upon Comparison of Norm and Criterion Referenced Assessment*, Master's Thesis, Ankara, 2013.

The institutions in education system use various assessment methods, while deciding the proficiency and the level of individuals in a particular content or construct. This correlational research model investigates that whether the decisions based on norm and criterion referenced assessment change by the type of assessment.

In the first phase of data collection, the multiple choice achievement test, composed of 20 questions measuring the level of the lesson called measurement and evaluation, was applied to 107 guidance and psychological counseling second grade students. The raw scores of each student were transformed into T scores, and they were classified into two groups as successful and unsuccessful by comparing each T score with the norm referenced criterion of 50 and 60 scores. Secondly, the opinion of 14 experts in measurement and evaluation was asked to apply Angoff and Nedelsky Methods. Thanks to the opinion of the experts, two different cut scores, in fact which are the criterion of success, were obtained by Angoff and Nedelsky Methods. The sum of raw scores obtained by achievement test was classified into two categories as successful and unsuccessful by comparing each raw score with the cut scores. In the final process of data collection, the students' semester scores in measurement and evaluation were obtained to study the concurrent validity of each method of norm and criterion referenced assessment.

The decisions based on the assessment methods were tested by z test, a statistical method measuring the meaning of difference between two dependent proportions. Afterwards, the concordance of the assessment methods with the semester scores was tested by Cohen's Kappa.

When the proportions of success based on the assessment methods were compared by z test, the results showed that each p-value was less than the value of alpha at 0,01. Accordingly, there are meaningful differences in the decisions based on norm and criterion referenced assessment. As for concurrent validity, Cohen's Kappa Test results showed that the concordance was only between semester scores and Nedelsky Method, which is a criterion referenced assessment method, and there was low concordance between semester scores and other assessment methods.

Key Words

Norm Referenced Assessment, Criterion Referenced Assessment, Angoff Method, Nedelsky Method

İÇİNDEKİLER

KABUL VE ONAY	i
BİLDİRİM	ii
TEŞEKKÜR.....	iii
ÖZET	iv
ABSTRACT.....	v
İÇİNDEKİLER	vi
TANIMLAR VE KISALTMALAR.....	viii
TABLolar DİZİNİ.....	ix
BÖLÜM I.....	1
GİRİŞ.....	1
1. 1. PROBLEM DURUMU	1
1. 1. 1. Norm Dayanaklı Değerlendirme (NDD).....	2
1. 1. 2. Ölçüt Dayanaklı Değerlendirme (ÖDD)	2
1. 1. 3. Norm Dayanaklı Değerlendirmelerle Ölçüt Dayanaklı Değerlendirmelerin Avantajları ve Dezavantajları.....	7
1. 1. 4. Norm Dayanaklı Değerlendirmelerle Ölçüt Dayanaklı Değerlendirmeler Arası Farklar	8
1. 1. 5. Norm Dayanaklı Değerlendirmelerle Ölçüt Dayanaklı Değerlendirmeler Arası Benzerlikler	9
1. 1. 6. Ölçme Aracında Bulunması Gereken Nitelikler	9
1. 2. PROBLEM CÜMLESİ	15
1. 3. SINIRLILIKLAR.....	16
1. 4. SAYILTILAR	16
1. 5. ARAŞTIRMANIN AMACI VE ÖNEMİ	16
1. 6. İLGİLİ ARAŞTIRMALAR	17
BÖLÜM II	22
YÖNTEM	22
2. 1. ARAŞTIRMANIN TÜRÜ	22
2. 2. ÇALIŞMA GRUBU.....	22
2. 3. VERİLERİN TOPLANMASI VE VERİ TOPLAMA ARACI	23
2. 4. VERİLERİN ANALİZİ	26
2. 4. 1. Birinci Problem.....	27

2. 4. 2. İkinci Problem.....	28
BÖLÜM III	30
BULGULAR VE YORUMLAR.....	30
3. 1. BİRİNCİ PROBLEME AİT BULGULAR VE YORUMLAR.....	30
3. 1. 1. Birinci Alt Probleme Ait Bulgular ve Yorumlar	33
3. 1. 2. İkinci Alt Probleme Ait Bulgular ve Yorumlar.....	34
3. 1. 3. Üçüncü Alt Probleme Ait Bulgular ve Yorumlar.....	35
3. 1. 4. Dördüncü Alt Probleme Ait Bulgular ve Yorumlar	36
3. 1. 5. Beşinci Alt Probleme Ait Bulgular ve Yorumlar	37
3. 2. İKİNCİ PROBLEME AİT BULGULAR VE YORUMLAR	38
3. 2. 1. Birinci Alt Probleme Ait Bulgular ve Yorumlar	38
3. 2. 2. İkinci Alt Probleme Ait Bulgular ve Yorumlar.....	39
3. 2. 3. Üçüncü Alt Probleme Ait Bulgular ve Yorumlar.....	40
3. 2. 4. Dördüncü Alt Probleme Ait Bulgular ve Yorumlar	41
BÖLÜM IV	43
SONUÇ VE ÖNERİLER.....	43
4. 1. SONUÇLAR.....	43
4. 1. 1. Birinci Probleme Ait Sonuçlar	43
4. 1. 2. İkinci Probleme Ait Sonuçlar.....	44
4. 2. ÖNERİLER.....	44
4. 2. 1. Araştırma Sonuçlarından Çıkan Öneriler.....	44
4. 2. 2. Araştırmacılara Yönelik Öneriler.....	45
KAYNAKÇA.....	47
EKLER.....	53
EK 1: Çalışmada Kullanılan Ölçme ve Değerlendirme Başarı Testinin Soruları	53
EK 2: Çalışmada Kullanılan Ölçme ve Değerlendirme Başarı Testinin Cevapları	61
EK 3: ITEMAN Sonuçlarına Göre Veri Toplama Aracına Seçilen Soruların Madde Güçlük ve Madde Ayırt Edicilik Gücü İndeksleri.....	62
EK 4: Angoff Yöntemi'ne Göre Test Maddelerinin Uzmanlarca Değerlendirilmesi	63
EK 5: Nedelsky Yöntemi'ne Göre Test Maddelerinin Uzmanlarca Değerlendirilmesi.....	64
EK 6: İzin Yazıları	65
ÖZGEÇMİŞ.....	66

TANIMLAR VE KISALTMALAR

Kesme Puanı: Ölçme ve değerlendirme başarı testinde yeterli seviyede olan kişilerle yeterli seviyede olmayan kişileri birbirinden ayıran performans noktasıdır.

KPSS: Kamu Personeli Seçme Sınavı

Minimum Yeterlik Seviyesi: “Eğitimle ilgili kararlar için uygun ölçme araçlarını seçebilen, gerektiğinde geliştirebilen, sonuçları puanlayabilen ve yorumlayabilen (değerlendirebilen), sonuçlardan faydalanabilen kişi” dikkate alındığında yeterli ile yetersiz arasındaki birey

NDD: Norm Dayanaklı Değerlendirme

ÖDD: Ölçüt Dayanaklı Değerlendirme

TABLOLAR DİZİNİ

Tablo 1: Kalan Seçenek Sayısına Göre Ulaşılabilecek Doğru Cevap Olasılıkları.....	6
Tablo 2: Araştırmaya Katılan Öğrenci ve Uzman Sayıları	22
Tablo 3: Veri Toplama Aracının Konu Başlıkları ve Soru Sayılarına Göre Dağılımı	24
Tablo 4: Nihai Uygulama Sonuçlarına Göre Testin Özellikleri.....	25
Tablo 5: Angoff ve Nedelsky Yöntemleri'ne Göre Uzmanlar Arası Uyum Katsayıları.....	30
Tablo 6: Angoff Yöntemi İçin Elde Edilen Kesme Puanlar	31
Tablo 7: Nedelsky Yöntemi İçin Elde Edilen Kesme Puanlar	31
Tablo 8: Yöntemlere Ait Kesme Puanları, Başarılı Kabul Edilen Öğrenci Sayısı ve Yüzdesi.....	32
Tablo 9: Angoff Yöntemi'ne ve 60 T Puanı Ölçütüne İlişkin Başarılı Kabul Edilen Öğrenci Sayıları-Yüzdeleri ve z Değeri.....	33
Tablo 10: Angoff Yöntemi'ne ve 50 T Puanı Ölçütüne İlişkin Başarılı Kabul Edilen Öğrenci Sayıları-Yüzdeleri ve z Değeri.....	34
Tablo 11: Nedelsky Yöntemi'ne ve 60 T Puanı Ölçütüne İlişkin Başarılı Kabul Edilen Öğrenci Sayıları-Yüzdeleri ve z Değeri.....	35
Tablo 12: Nedelsky Yöntemi'ne ve 50 T Puanı Ölçütüne İlişkin Başarılı Kabul Edilen Öğrenci Sayıları-Yüzdeleri ve z Değeri.....	36
Tablo 13: 50 ve 60 T Puanı Ölçütlerine İlişkin Başarılı Kabul Edilen Öğrenci Sayıları-Yüzdeleri ve z Değeri	37
Tablo 14: 60 T Puanı Ölçütüne İlişkin Başarı Durumu ve Dönem Sonu Ders Başarısı Durumu Arasındaki Uyum Frekansı, Yüzdesi ve Kappa Değeri	38
Tablo 15: 50 T Puanı Ölçütüne İlişkin Başarı Durumu ve Dönem Sonu Ders Başarısı Durumu Arasındaki Uyum Frekansı, Yüzdesi ve Kappa Değeri	39
Tablo 16: Angoff Yöntemi'ne İlişkin Başarı Durumu ve Dönem Sonu Ders Başarısı Durumu Arasındaki Uyum Frekansı, Yüzdesi ve Kappa Değeri	40
Tablo 17: Nedelsky Yöntemi'ne İlişkin Başarı Durumu ve Dönem Sonu Ders Başarısı Durumu Arasındaki Uyum Frekansı, Yüzdesi ve Kappa Değeri	41

BÖLÜM I

GİRİŞ

Bilimsel araştırma; karşılaşılan sorunu tanımlama, soruna ilişkin hipotezler kurma ve bu hipotezleri sınama sürecinden oluşmaktadır (Yıldırım, 2011). Bu doğrultuda karşılaşılan sorunun tanımlanacağı bu bölümde sorunun ne olduğu, nedenlerinin neler olduğu, soruna ilişkin eski çalışmaların nasıl sonuçlandığı ve sorunun neden çalışmaya değer olduğu açıklanmıştır.

1. 1. PROBLEM DURUMU

Geçmişten günümüze hem özel sektörde hem de devlet kurumlarında nitelikli bireylerin seçimi veya bireylerin eğitimleri hep önemini korumuştur. Kurumlar misyonları ve vizyonları doğrultusunda daha başarılı işler çıkarmak için eleman seçimine önem vermiş ve seçilen elemanların çeşitli eğitimlerle niteliklerini artırmaya çalışmışlardır. Günümüzde ise kurumlar ihtiyaç duydukları nitelikli bireyleri belirlerken veya verilen eğitimler doğrultusunda bireylerin istenilen hedeflere ne ölçüde ulaştıklarını tespit ederken çeşitli yöntemler kullanmaktadır. Bu noktada kullanılan yöntemlerin uygun olması gerekmektedir. Aksi takdirde hem bireyler hem de kurumlar çeşitli zararlar görebilmektedir. Yanlış bireyleri seçen kurumlar niteliksiz iş gücü nedeniyle ekonomik zararlara uğrarken kendisine uygun olmayan işlerde çalışan bireyler psikolojik anlamda zarar görmekte ve yıpranmaktadır.

Eğitim kurumları, bireylerin seçimi ve bireylerin verilen eğitimlerle hedeflere ne ölçüde ulaştıklarını belirlemede önemli kararlar verebilmektedir. Bu kararların alınmasında kurumların oldukça dikkatli olmaları gerekmektedir. Alınan yanlış kararlar bir sonraki hedefe ulaşma noktasında sorunlara neden olabilmektedir. Söz konusu yanlış kararları en aza indirebilmek için uygun karar verme mekanizmaları geliştirmek gerekmektedir. Bu noktada uygun ölçme ve değerlendirme yöntemlerini kullanmak önemlidir.

Ölçme ve değerlendirme birbiriyle iç içe olan; ancak aynı zamanda farklı tanımları olan kavramlardır. Ölçme niteliklerin sayı ve sembollerle eşlenmesidir (Turgut ve Baykul, 2010). Ölçme bize sadece bir niteliğin niceliğine ilişkin bilgi verir. Ancak bu bilginin yeterli olup olmadığına ilişkin bilgiyi değerlendirme verir. Bu doğrultuda bir ölçme sonucu tek başına yeterli değildir ve ölçme sonuçlarının bir ölçütle kıyaslanarak değerlendirilmesi gerekmektedir.

Değerlendirme bir karar verme işidir ve ölçme sonuçlarının bir ölçütle karşılaştırılarak karara varılması sürecini kapsar. Yani değerlendirme; ölçme sonuçları, ölçüt ve karar olmak üzere üç ögeyi içerir. Bu doğrultuda değerlendirme, ölçme sonuçlarını bir ölçütle karşılaştırarak ölçülecek nitelik hakkında karar vermeye olanak sağlar. Bu nedenle ölçme sonuçlarını değerlendirebilmek için ölçütün tanımlanması gerekir (Baykul, 2010).

Turgut ve Baykul'a (2010) göre ölçüt, ilgili değerlendirme için ölçme sonuçlarının karşılaştırıldığı, değerlendirme için değişmeyen değerlerdir. Gözlem sonuçları ölçüte göre yorumlanır ve bir yargıya ya da karara varılır. Değerlendirme sürecinde kullanılan ölçütlere göre değerlendirmeyi, norm dayanaklı değerlendirme (NDD) ve ölçüt dayanaklı değerlendirme (ÖDD) olmak üzere iki sınıfta tanımlayabiliriz (Murphy ve Davidshofer, 2005).

1. 1. 1. Norm Dayanaklı Değerlendirme (NDD)

Norm dayanaklı değerlendirmeler bireyin performansı ile grup performansını karşılaştırmayı sağlayan formal değerlendirmelerdir. NDD, bireyleri ya da nesnelere birbirleriyle karşılaştırarak değerlendirme yapmayı amaçlamaktadır. Bu doğrultuda grubun istatistiklerine göre bireyi ya da nesneyi grup ile karşılaştırır ve bağıl ölçüt kullanır. Bu değerlendirmelerin açık bir yönergesi ve puanlama ölçütü vardır. Bu sayede bu tarz değerlendirmelerin geçerliğini, güvenilirliğini, ölçme hatasını ve değişkenliğini değerlendirmek kolaylaşmaktadır. NDD, standartlaştırılmış değerlendirmeler olup ölçmenin kapsadığı her birey ya da nesne aynı süreçlerden geçmektedir (Paul, 1995).

Norm dayanaklı değerlendirmeler iki farklı şekilde olabilmektedir. İlk değerlendirmede gözlem yapılan grup dikkate alınarak bireyler kendi aralarında karşılaştırılır. İkinci tür değerlendirmede ise gözlem yapılan davranışlarla ilgili bir referans grup belirlenir ve yeni yapılan gözlemlerde bireyler söz konusu referans grup ile karşılaştırılır. Özellikle ikinci tür değerlendirmelerde norma bağlı doğru değerlendirmeler yapabilmek için doğru referans grubunun seçimi önem kazanmaktadır (Urbina, 2004).

1. 1. 2. Ölçüt Dayanaklı Değerlendirme (ÖDD)

Ölçüt dayanaklı değerlendirmelerde birey ya da nesne diğerleriyle karşılaştırılmak yerine belirli bir performans seviyesi ile karşılaştırılmaktadır. Bu doğrultuda bireyin performansı belirli bir bilgi alanı ya da yetenek seti ile karşılaştırılmaktadır. Bu nedenle ÖDD bireyin performansındaki değişimi ölçmek için tasarlanmaktadır. Mutlak ölçütün kullanıldığı ÖDD, özellikle program planlama ve değerlendirme sürecinde kullanılmaktadır (Paul, 1995). Ölçüt

dayanaklı deęerlendirmelerde ölçüt belirlenirken kullanılan yöntemlerden birisi standart belirleme yöntemleri ile uzman görüşüne göre hesaplanan kesme puanlarıdır (Urbina, 2004).

1. 1. 2. 1. Standart Belirleme

Standart belirleme performans seviyelerini ayırmaya yarayan kesme puanını hesaplama sürecidir. Uzman görüşüne dayanan yöntemde performans düzeyine ilişkin kesme puanı hesaplanır ve hesaplanan puan ölçme sonuçlarının deęerlendirilmesinde ölçüt olarak kullanılır. Standart belirleme yöntemleriyle kesme puanı hesaplanırken “Ne kadar yeterlidir?” sorusuna cevap aranmaktadır. Bu doğrultuda kesme puanı, testin amacı açısından yeterli bilgi veya yetenek seviyesini göstermektedir (Livingston ve Zieky, 1989). Kesme puanına ilişkin yeterli bilgi ve yetenek seviyesini gösteren performans tanımlanırken çoğunlukla alan uzmanlarının görüşleri dikkate alınmaktadır (Margolis, 2011).

Kesme puanı hesaplamak için görüşüne başvurulacak uzmanların seçimi, standart belirleme sürecindeki önemli bir adımdır; çünkü yanlış uzmanların seçilmesi nedeniyle yanlış bir kesme puanına ve yanlış kararlara ulaşılabilir. Bu noktada testin ve uygulamanın amacı doğrultusunda uzmanların seçimi yapılmaktadır. Genel olarak uzman seçiminde uygulamanın konusu açısından deneyimli olmak dikkate alınmaktadır. Bunun yanında uygulamanın yapılacağı grubun tanınması da uzman seçiminde önemli bir etken olmaktadır. Seçilen uzmanların yeterli sayıda olması da önemlidir (Smith, 2011).

Performansa ilişkin sınıflama yapmaya yarayan standart belirleme yöntemleri pek çok şekilde hesaplanmaktadır. Jaeger (1989), standart belirleme yöntemlerini test merkezli ve öğrenci merkezli olmak üzere iki sınıfta toplamıştır.

1. 1. 2. 1. 1. Test Merkezli Standart Belirleme Yöntemleri

Test merkezli yöntemler, testte yer alan maddelere ilişkin uzman görüşüne dayanmaktadır. Uzmanlar belirlenen performans düzeyine göre her bir maddeyi deęerlendirir ve bu süreç sonunda kesme puanı hesaplanır (Aktaran: Taşdelen, 2009, s. 5). Angoff, Ebel, Nedelsky, Yes/No ve İşaretleme yöntemleri test merkezli standart belirleme yöntemlerindedir.

1. 1. 2. 1. 1. 1. Angoff Yöntemi

Angoff Yöntemi, kullanışlı standart belirleme yöntemlerinden biridir. Uygulaması, veri toplaması ve verileri analiz edilmesi açısından kolay olması nedeniyle sıkça kullanılan bir

yöntemdir. Özellikle çoktan seçmeli testler için uygundur. Bu doğrultuda çoktan seçmeli testlerin sıkça kullanıldığı lisans ve sertifika programlarında yaygın olarak kullanılmaktadır (Cizek ve Bunch, 2007).

Angoff Yöntemi'ne göre kesme puanı hesaplanırken ilk olarak ölçülen özellik bakımından minimum yeterli düzeyi tanımlanır (Livingston ve Zieky, 1989). Bu tanım dikkate alınarak konu alanı uzmanları, her bir maddenin minimum yeterli düzeyindeki kişiler tarafından cevaplanabilme olasılıklarını belirler. Elde edilen olasılıklar toplanarak uzman sayısına göre ortalaması alınır. Nihai puan, testin kesme puanını oluşturur. Hesaplama formülü şu şekilde ifade edilmiştir (Aktaran: Tanrıverdi, 2006, s. 11):

$$T_{cj} = \sum_{i=1}^n (p_{ij})$$

T_{cj} : j uzmanın belirlediği kesme puanı

p_{ij} : minimum yeterli düzeyindeki bir öğrencinin j uzmanı tarafından belirlenen i maddesini doğru cevaplayabilme olasılığı

n : testteki madde sayısı

1. 1. 2. 1. 1. 2. Ebel Yöntemi

Ebel Yöntemi'nde uzmanlar her bir maddeyi iki boyut üzerinden değerlendirir. Bunlar her bir test maddesinin testte ölçülen özelliğe uygunluğu ve zorluğudur. Uzmanlar her bir test maddesini üç satır ve dört sütundan oluşan bir tabloya yerleştirirler. Tablonun satır kısmı testin kolay, orta ve zor olmak üzere zorluk boyutunu; sütun kısmı ise gerekli, önemli, kabul edilebilir ve tartışılabilir olmak üzere uygunluk boyutunu gösterir. Uzmanlar ayrıca ilgili hücrelerde, minimum yeterli seviyesindeki öğrencinin doğru cevaplayabileceği madde yüzdelerini de belirtir. Hücrelerdeki test maddeleri toplanarak uzmanların belirttiği yüzdelere çarpıldıktan sonra tüm hücrelerden elde edilen puanlar toplanır. Uzmanlara ilişkin elde edilen kesme puanlarının ortalaması, testin kesme puanını verir (Alsmadi, 2007).

1. 1. 2. 1. 1. 3. Nedelsky Yöntemi

Nedelsky tarafından 1954 yılında önerilen yöntemin günümüzde de kullanılmaya devam edilmesinde, uygulama kolaylığı önemli bir sebeptir. Yöntemin ortaya atıldığı yıllarda norm dayanaklı değerlendirmeler daha yaygın olarak kullanılıyordu. Nedelsky Yöntemi, kişilerin kendi aralarında karşılaştırılmasına dayalı performans seviyelerinden, kişilerin gerçek performans seviyelerini gösteren değerlendirme biçimine geçişi hızlandıran yöntemlerden birisidir (Cizek ve Bunch, 2007).

Angoff Yöntemi gibi Nedelsky Yöntemi de test maddeleri üzerinden kesme puanı hesaplamaya çalışır. Bu yöntemi kullanarak kesme puanı hesaplamada her bir maddedeki seçenek sayısı önemlidir (Arrasmith, 1986). Testte yer alan maddelerin yanlış seçenekleri üzerinden giderek test maddesinin çözülmesine ilişkin bir olasılık hesaplanır.

Uzmanlar testte ölçülen özelliği dikkate alarak minimum yeterlikteki kişinin her bir test maddesinde yer alan kaç yanlış seçeneği doğru tahmin edebileceklerine karar verirler. Bu sayede her bir test maddesi için minimum yeterlikteki kişiler tarafından elenemeyecek seçenek sayısı bulunur. Elenemeyen seçenekler arasından doğru cevabı bulma olasılığı ise madde puanını verir. Örneğin bir uzman, beş seçenekli bir maddedeki seçeneklerden ikisinin minimum yeterlikteki kişi tarafından eleneceğini tahmin etmektedir. Bu durumda geriye 3 seçenek kalmaktadır. Bir kişinin kalan üç seçenektan doğru cevabı bulma olasılığı ise üçte birdir. Bu doğrultuda söz konusu maddenin uzman tarafından tahmin edilen çözülebileme olasılığı 0.33 olarak hesaplanır. Hesaplanan tüm olasılıklar ise toplanarak her bir uzmana ilişkin kesme puanı bulunur. Kesme puanlarının aritmetik ortalaması alınarak Nedelsky Yöntemi'ne ilişkin testin kesme puanı hesaplanmış olur (Shin, 2004). Hesaplama formülü aşağıdaki gibi ifade edilmiştir (Aktaran: Tanrıverdi, 2006, s. 9).

$$T_{cj} = \sum_{i=1}^n (q_i - k_{ij})^{-1}$$

T_{cj} : j uzmanın belirlediği kesme puanı

q_i : i maddesindeki seçenek sayısı

k_{ij} : j uzmanının minimum yeterlikteki bir öğrencinin yanlış olarak elenebileceğini düşündüğü seçenek sayısı

n : testteki madde sayısı

Nedelsky Yöntemi uygulandığında, beş seçenekli bir test için uzman görüşü dahilinde kalan seçenek sayısına göre her bir maddeye ilişkin ulaşılabilir olasılıklar Tablo 1’de yer almaktadır.

Tablo 1: Kalan Seçenek Sayısına Göre Ulaşılabilir Doğru Cevap Olasılıkları

Kalan Seçenek Sayısı	Doğru Cevap Olasılığı
1	1.00
2	.50
3	.33
4	.25
5	.20

Seçenek sayısı arttığında doğru cevap olasılığı düşmektedir. Öğrencinin gelişim düzeyi veya okul düzeyine göre dört ya da beş seçenekli çoktan seçmeli test maddeleri kullanılmaktadır. Bu durumda zor çeldiricilerin olduğu test maddeleri için elenebilen seçenek sayısı az olmakta ve çoğunlukla doğru cevap olasılığı .20 ve .25 olarak karşımıza çıkmaktadır.

1. 1. 2. 1. 1. 4. Yes/No Yöntemi

Yes/No Yöntemi, Impara ve Plake (1997) tarafından Angoff Yöntemi’nin bir çeşitlemesi olarak geliştirilmiştir. Bu yöntemde uzmanlar, testte ölçülen özellik bakımından minimum yeterlikte olan kişilerin her bir test maddesini doğru cevaplayıp cevaplayamayacaklarını tahmin etmeleri istenir. Uzmanların evet cevabı, 1 puan; hayır cevabı, 0 puan olarak kodlanır. Uzmanlar bu şekilde maddeleri puanlar ve puanlanan maddelerin toplamı, söz konusu uzmanın teste ilişkin kesme puanını verir. Tüm uzmanların puanlarının ortalaması alınarak testin kesme puanı hesaplanır (Cizek ve Bunch, 2007).

1. 1. 2. 1. 1. 5. İşaretleme Yöntemi

İşaretleme Yöntemi’nde diğer yöntemlerden farklı olarak her bir test maddesini dikkate almak yerine testin bütünü dikkate alınmaktadır. Madde tepki kuramına göre test maddeleri kolaydan zora doğru sıralanır ve bu sıralama dikkate alınarak uzmanlar tarafından kesme puanının sınırı “İleri Düzey”, “Yeterli” ve “Yetersiz” olmak üzere üç grup için belirlenir. Bu sayede maddeler,

performans seviyesi bakımından birbirinden ayrılacak şekilde işaretlenir (Gelbal ve Çetin, 2010).

1. 1. 2. 1. 2. Öğrenci Merkezli Standart Belirleme Yöntemleri

Öğrenci merkezli yöntemlerde, öğrenciler uzmanlar tarafından sınıflandırılır. Sınıflandırılmaya uygun olacak şekilde bir kesme puanı belirlenir. Öğrenci merkezli standart belirleme yöntemlerinin en yaygın olarak kullanılanları sınır grup yöntemi ve karşıt gruplar yöntemidir (Kane, 1998).

1. 1. 2. 1. 2. 1. Sınır Grup Yöntemi

Livingstone ve Zieky tarafından geliştirilen bu yöntemde testi alan kişileri tanıyan uzmanlar, testin içeriğine göre öğrencileri “Yeterli”, “Sınırdan” ve “Yetersiz” olmak üzere üç gruba ayırır. Bu süreçten sonra test uygulanır ve uzmanlar test puanlarını dikkate alarak yeterli ve yetersiz grupta yer alan öğrencilerden hangilerinin söz konusu iki grubun arasında, sınırdan olduğunu belirlerler. Sınır grup olarak belirlenen öğrencilerin puanlarının ortancası ise kesme puanını verir (Livingston ve Zieky, 1989).

1. 1. 2. 1. 2. 2. Karşıt Gruplar Yöntemi

Karşıt gruplar yönteminde testi alan kişiler konu alanı uzmanları tarafından “Yeterli” ve “Yetersiz” olarak iki gruba ayrılır. Daha sonra test uygulanır ve hatalı kararları en aza indirecek şekilde çeşitli yöntemlerle kesme puanı hesaplanır. Söz konusu hesaplamada mümkün olduğu kadar çok sayıda, yeterli kabul edilen kişilerin kesme puanı üzerinde puan almaları ve yetersiz kabul edilen kişilerin de kesme puanı altında puan almaları dikkate alınır (Norcini, 2003).

1. 1. 3. Norm Dayanaklı Değerlendirmelerle Ölçüt Dayanaklı Değerlendirmelerin Avantajları ve Dezavantajları

Norm dayanaklı değerlendirmelerin çeşitli avantajları ve dezavantajları vardır. NDD ile bireylerin performanslarının ortalamadan sapmalarına göre bireyler sıralanır. Bu sıralama bireyler arası farklılıkları ortaya koyar. Bu noktada bireyin performansını sınıflandırma veya bireyi belirli bir performans seviyesinde tanımlama gücü NDD için bir dezavantaj olmaktadır; çünkü norm dayanaklı değerlendirmede bireylerin aldıkları puanlar, ölçülen özellik bakımından gerçek seviyeyi ve puanı göstermemektedir (Johnson ve Martin, 1980).

Norm dayanaklı deęerlendirmelerin bir başka dezavantajı ise söz konusu deęerlendirmelerde elde edilen puanlar, bireyleri karşılaştırmaya dayalı tahmin edilen puanlardır. Bu nedenle bireylerin seviyelerini, belirli bir özellikte kategorilerini ve bir programın ne kadar etkili olduğunu deęerlendirmede yanlış kullanımlara açıktır (McCauley ve Swisher, 1984).

Norm dayanaklı deęerlendirmelerin avantajları dikkate alınırca, McCauley ve Swisher (1984) norm dayanaklı deęerlendirmelerin seçme ve yerleştirme süreçlerinde kullanılması açısından avantajlı olduğunu belirtmektedir. Deęerlendirmeler sonucunda belirli sayıda kişinin seçileceęi süreçlerde norm dayanaklı deęerlendirmeler kullanışlıdır. Bireyler arası farklılıkları dikkate alarak bireyleri ölçülen özellik bakımından sıralayan norm dayanaklı deęerlendirmeler yapılarak seçilecek kişiler, grup sıralamasında en yüksek puan alanlar olarak belirlenebilir.

Ölçüt dayanaklı deęerlendirmelerin de bazı avantajları ve dezavantajları vardır. Bu tarz deęerlendirmeler özellikle öğrenmenin etkilerini ölçme anlamında önemlidir. Bu noktada test ile program hedefleri doğrudan ilişkilendirilebilmektedir. Ayrıca ölçüt dayanaklı deęerlendirmeler bireylerin yeteneklerini ve ihtiyaçlarını ölçme bakımından son derece kullanışlıdır. Yetenekler ve ihtiyaçlar ölçüt dayanaklı deęerlendirmelerle ölçülerek öğrenme süreci yeniden planlanabilir (Freeman ve Miller, 2001).

Ölçüt dayanaklı deęerlendirmelerin avantajlarının yanında bazı dezavantajları da vardır. Ölçüt dayanaklı deęerlendirmelerde eęer normlar tanımlanmamışsa çeşitli yaş gruplarına uygulama güçlüğü vardır. Ölçüt dayanaklı deęerlendirmelerin bir başka dezavantajı ise kişilerin karşılaştırılarak yorumlanması güçlüğüdür (MacTurk ve Neisworth, 1978).

1. 1. 4. Norm Dayanaklı Deęerlendirmelerle Ölçüt Dayanaklı Deęerlendirmeler Arası Farklar

NDD ile ÖDD arasındaki en önemli farklardan birisi, iki deęerlendirme yönteminin kullanım amaçlarının farklı olmasıdır. Norm dayanaklı deęerlendirmede amaç bireyleri sıralamak, ölçüt dayanaklı deęerlendirmede ise amaç kişilerin performans seviyelerini belirlemektir. Bu doğrultuda performans deęerlendirilirken norm dayanaklı deęerlendirmelerde yüzdesel sıralama ya da standart puanlar kullanılırken, ölçüt dayanaklı deęerlendirmelerde ham puan dikkate alınmaktadır (McCauley, 1996).

İki farklı deęerlendirme açısından herhangi bir kişinin bir testten aldığı puan deęerlendirilirken farklı yaklaşımlarda bulunmaktadır. Norm dayanaklı deęerlendirmelerde bireylerin testten aldıkları puanlar, birbirleriyle karşılaştırılarak baęlı deęerlendirme yapılmaktadır. Kişinin teste ilişkin seviyesi grubun neresinde olduğuna baęlıdır. Ölçüt dayanaklı deęerlendirmelerde ise

belirlenen geme puanına gre kiřinin bařarı sı diđer bireylerle karřılařtırılmadan deęerlendirilir. Bu tarz deęerlendirmelerde ise mutlak deęerlendirme yapılır. Bařarı sı, minimum bařarı seviyesinde ya da ok bařarılı gibi seviyeler doęrultusunda belirlenen kesme puanlarına gre bireyin performansı deęerlendirilir (Mehrens ve Lehmann, 1991).

NDD ve DD arasındaki farklardan bir diđer i se iki deęerlendirme řeklinin odaęındaki unsurlardır. NDD odaęında insan unsuru daha ok n planda yer alırken, DD odaęında bir kapsam alanına iliřkin bilgi seti ya da performans kalitesine dayalı yetkinlik seviyesi bulunmaktadır. Deęerlendirmeler yapılıırken sz konusu unsurlar dikkate alınmakta ve bu doęrultuda kararlar verilmektedir (Urbina, 2004).

İki farklı deęerlendirmenin kullanıldıęı testlerin zellikleri aısından da farklılık vardır. zellikle yapıların ve ierięin iyi tanımlandıęı durumlarda iki deęerlendirme řekli de kullanılabilirken, sz konusu yapıların ve ierięin yeterince tanımlanamadıęı durumlarda lt dayanaklı deęerlendirmeler yapmak gleřmektedir. rneęin zeka, davranıř boyutunda yapı ve ierik olarak yeterince tanımlanamadıęı iin lt dayanaklı deęerlendirme yapmak yeterince uygun olmamaktadır (Mehrens ve Lehmann, 1991).

1. 1. 5. Norm Dayanaklı Deęerlendirmelerle lt Dayanaklı Deęerlendirmeler Arası Benzerlikler

NDD ile DD arasında birok fark olmasına raęmen bazı benzer ynler de bulunmaktadır. İki deęerlendirme řeklinin kullanım amacı aısından karřılařtırılması yapılırsa, geerli kararlar vermede ikisinin de kullanılması gerektięi durumlar vardır. Kiřilere rehberlik etmek amacıyla kullanılan testlerde iki farklı deęerlendirme řekli de uygulanabilir. Bunun yanında kiřileri bařarılarına ya da seviyelerine gre sınıflandırmada hem norm dayanaklı deęerlendirme hem de lt dayanaklı deęerlendirme uygun olabilmektedir (Mehrens ve Lehmann, 1991).

İki deęerlendirme řekli arasındaki bir bařka ortak nokta, her iki deęerlendirmenin de yapıldıęı testlerin uygun psikometrik zelliklere sahip olması gereklilięidir. Bu doęrultuda kullanılan testlerin mutlaka gvenilir ve geerli olması řarttır. Bu nedenle norm dayanaklı bir test ile lt dayanaklı bir testin uygulanması iin her iki testin de kapsam, yapı ve lt gibi geerliklerinin saęlanması nemlidir (Montgomery ve Connolly, 1987).

1. 1. 6. lme Aracında Bulunması Gereken Nitelikler

Kiřiler ya da nesnelere hakkında deęerlendirme yaparken ve kararlar alırken eřitli gzlemler yapılmaktadır. Sz konusu kiřilere ya da nesnelere ait ęrenmek istenilen zellikler, gzlemler

sayesinde ölçülmeye çalışılır ve gözlemlerin bir amacı vardır. Bu amaç doğrultusunda ölçmeye çalıştığımız söz konusu özelliklere ilişkin nihai bir karara varabilmek için farklı kişiler tarafından veya farklı zamanlarda yapılan gözlemlerin ise tutarlı olması gerekmektedir. Ancak gözlemler çoğu zaman doğrudan değil, çeşitli ölçme araçları kullanılarak dolaylı olarak yapılmaktadır. Bu doğrultuda kullanılan ölçme araçlarının gözlem yapma amacına uygun olması ve tutarlı sonuçlar vermesi gerekmektedir. Güvenirlik ve geçerlik kavramları ise bu noktada gündeme gelmektedir (Muijs, 2004). Söz konusu güvenilirlik ve geçerlik kavramları bir ölçme aracında bulunması gereken en önemli özelliklerdir ve herhangi biri diğerine tercih edilemez (Baykul, 2010).

1. 1. 6. 1. Güvenirlik

Turgut ve Baykul'a (2010) göre güvenilirlik, ölçme sonuçlarının tesadüfi hatalardan arınıklık derecesidir. Ölçme hatasının kestirilebilmesi ise ölçmenin tekrarı sayesinde mümkündür. Bu doğrultuda psikometrik anlamda güvenilirlik kelimesi, ölçme tutarlılığı anlamına gelmektedir ve bir ölçme aracını kullanabilmek için söz konusu ölçme aracının güvenilir olması gerekmektedir (Domino ve Domino, 2006). Bu noktada dikkat çekilen güvenilirlik, ölçme aracının tamamen güvenilir olması ya da hiç güvenilir olmaması anlamında değildir; çünkü bir ölçme aracı bazı özellikler açısından güvenilir olabilirken başka özellikler açısından güvenilir olamayabilmektedir (Cohen ve Swerdlik, 2009).

Söz konusu özellikler açısından bir testin güvenilirliği test tekrar test, eşdeğer formlar, iç tutarlılık ve eşdeğer yarılar yöntemleriyle belirlenebilmektedir. Bu yöntemler ile elde edilen güvenilirlik indeksi 0,00 ile +1,00 arasında değişen değerler almaktadır. Güvenirlik indeksinin +1,00 olması mükemmel güvenilirlik anlamına gelirken 0,00 olması ise tamamıyla hatalardan oluşan ölçmeyi ifade etmektedir (Turgut ve Baykul, 2010).

1. 1. 6. 2. Geçerlik

Bazı durumlarda çok yüksek güvenilirlik, testin geçerliğinin düşmesine neden olsa da bir testin geçerli olması için güvenilir olması şarttır. Ancak güvenilirlik, geçerlik için tek başına yeterli değildir (Kline, 2000). Çünkü bir özelliği ölçmek için geliştirilen güvenilir bir ölçme aracı aslında başka bir özelliği ölçüyor olabilir ve güvenilir olmasına rağmen araştırmacının amacına hizmet etmiyor olabilir. Bu nedenle güvenilir bir ölçme aracı olmasına rağmen geçerliği düşük olabilir (Murphy ve Davidshofer, 2005). Bu doğrultuda bir ölçme aracının amacına uygun sonuçlar verip vermediğinin incelenmesi gerekmektedir.

Geçerlik, bir ölçme aracının ölçmek için tasarlandığı özelliği ya da değişkeni ölçme düzeyidir (Kline, 2000; Aiken, 2000; Baykul, 2010; Turgut ve Baykul, 2010). Bu tanım dikkate alındığında şaşırtıcı bir şekilde birçok ölçme aracının düşük geçerliğe sahip olduğu görülmektedir. Bu durum bir ölçme aracının hiçbir yönden geçerli olmadığı ve kullanılamayacağı anlamına gelmemektedir; çünkü bir ölçme aracının bazı amaçlar açısından mutlaka geçerli olduğu yönleri vardır ve bazı açılardan diğerlerine oranla daha geçerlidir. Bunun nedeni ise bir teste ait tek bir geçerlik katsayısından bahsedilememesindedir. Bu doğrultuda bir testin geçerliği saptanırken kullanılan farklı yöntemler vardır (Kline, 2000). Kapsam ve yapı geçerliği yöntemleri ile ölçmenin geçerliği; yordama ve uygunluk geçerlikleri ile kararların geçerliği incelenebilmektedir (Murphy ve Davidshofer, 2005).

Kapsam geçerliği, bir testin geçerliğini saptarken kullanılan yöntemlerden biridir ve bir ölçme aracıyla ölçülmek istenilen özelliklere ilişkin davranışları söz konusu ölçme aracının ne seviyede içerdiği şeklinde tanımlanmaktadır (Domino ve Domino, 2006). Ancak bir testin ölçmek istenilen kapsama ilişkin tüm davranışlara yönelik maddeler içermesi çoğunlukla mümkün olmamakta ve genellikle söz konusu kapsamı temsil eden maddeler oluşturularak örnekleme gidilmektedir. Bu noktada içerik ve davranış boyutları çerçevesinde hazırlanan belirtke tablolarından faydalanılmaktadır. Belirtke tabloları ile hazırlanan maddelerin ölçülmek istenilen kapsam alanını ne kadar temsil ettiği ise kapsam geçerliğini belirlemektedir. Özellikle başarı testlerinde kapsam geçerliğine başvurulmaktadır (Aiken, 2000).

Yapı geçerliği, bir ölçme aracıyla ölçülmek istenilen yapının söz konusu ölçme aracı tarafından ortaya konulma seviyesi olarak tanımlanmaktadır (Baykul, 2010). Ölçme aracının ölçülmek istenilen yapıyı doğru şekilde ölçüp ölçmediğiyle ilgilidir (Murphy ve Davidshofer, 2005). Çeşitli yöntemlerle yapı geçerliği incelenebilmektedir. Testte ölçülen özellik ile aynı özelliği ölçen diğer testlerin korelasyonuna bakılabilir. Ayrıca testte ölçülen özellik ile benzerlik gösteren başka yapıları ölçen test sonuçları arasındaki korelasyon da incelenebilir. Testte ölçülen yapı açısından farklı özelliklere sahip olduğu düşünülen iki grup puanları arasındaki korelasyon alınarak da testin yapı geçerliği incelenebilir. Araştırmacılar tarafından çok sık kullanılan bir başka yöntem ise faktör analizidir (Baykul, 2010).

Bir ölçüte dayalı geçerlik, ölçme sonuçları ve belirlenen ölçütler arasındaki ilişki şeklinde tanımlanmaktadır (Coalay, 2010). Bir başka deyişle yordayan ve yordanan arasındaki ilişki, bir ölçüte dayalı geçerlik hakkında bize bilgi vermektedir. Bu bilgi sayesinde gelecekle ilgili tahminlerde bulunulabilmekte ve bireyler hakkında kararlar verilebilmektedir. Bu nedenle insanlar hakkında karar verebilmek için ölçme aracının söz konusu geçerliğe sahip olması son

derece önemlidir. Ölçüte ilişkin ölçme sonuçlarının elde edildiği zamana bağlı olarak iki farklı ölçüte dayalı geçerlik şekline söz edilebilir (Muijs, 2004).

Ölçüte dayalı geçerliklerden ilki olan yordama geçerliği, ölçme sonuçlarının söz konusu özellikle ilişkili gelecekteki bir özelliği kestirebilmesi şeklinde tanımlanmaktadır (Kline, 2000). Burada dikkate alınan ölçüt, ölçme sonuçlarından sonra, gelecekte elde edilmektedir. Eğitim sistemi dikkate alındığında ölçme sonuçları öğrencinin başarısına ilişkin kararlarda kullanılmakta ve başarıya göre öğrencilere not verilmektedir. Not verme ise bir sınıflama işlemidir ve bu doğrultuda öğrenciler “Geçti” ve “Kaldı” şeklinde sınıflanmaktadır (Turgut ve Baykul, 2010). Yordama geçerliği kapsamında söz konusu kararların bir dış ölçütle ilişkili olması ise sınıflamaya dayalı kararların geçerliğini vermektedir (Murphy ve Davidshofer, 2005; Erkuş, 2004).

Yordama geçerliğinde özellikleri ölçülen kişilerden bazıları alınan kararlar ile çalışma dışında tutulabilmektedir. Bu nedenle testte ölçülen yapı ile ilişkili gelecekteki bir başka yapı yordanırken, çalışma dışında tutulan kişilerin verileri dikkate alınmadığı için tam anlamıyla geçerli sonuçlara ulaşamamaktadır. Ölçüte dayalı geçerliklerden ikincisi olan uygunluk geçerliği ile bu dezavantajın önüne geçilmektedir. Uygunluk geçerliği ile ölçülen özellik, gelecekteki başka bir özelliği değil, mevcut bir özelliği ya da geçmişte elde edilen bir özelliği ne kadar kestirdiği dikkate alınır (Marshall ve Hales, 1971).

Sonuç olarak ölçüte dayalı geçerlik türleri sayesinde, ölçme sonuçları doğrultusunda alınan kararların ne kadar başarılı ve geçerli olduğu incelenebilir. Bu nedenle kişiler hakkında karar almanın ön planda olduğu ölçme ve değerlendirme süreçlerinde ölçüte dayalı geçerliği dikkate almak önemlidir. Ölçme sonuçları ile karşılaştırılan ölçüt arasındaki yüksek ilişki, alınan kararların başarısı ve geçerliği hakkında bilgi verecektir.

Ölçme ve değerlendirme süreçleri sonucunda alınan kararların başarılı olabilmesi için ölçme aracının güvenilir ve geçerli olmasının yanı sıra kullanılan değerlendirme yönteminin de uygun olması gerekmektedir. İki farklı değerlendirme yöntemi olan norm dayanaklı değerlendirmelerle ölçüt dayanaklı değerlendirmelerin avantajları, dezavantajları ve farkları dikkate alındığında doğru karar verebilmek için hangi durumda hangi yöntemin kullanılacağını bilmek önem kazanmaktadır.

Mevcut uygulamalarda kullanılan değerlendirme yöntemleri incelendiğinde benzer durumlar için farklı yöntemler uygulandığı görülmektedir. Örneğin milli eğitim sisteminde tüm kademelerde amaç belirli hedefler doğrultusunda, belirli bir eğitim programı dahilinde öğrencilerde davranış değişikliği kazandırmaktır. Ancak yükseköğretime kadar olan

kademelerde ölçüt dayanaklı değerlendirmeler kullanılırken yükseköğretimde kullanılan yöntemler üniversiteden üniversiteye değişmektedir.

Ülkemizde eğitim sistemi çerçevesinde kurumlar, öğrenci başarılarını çeşitli yönetmeliklere göre değerlendirmektedir. Milli Eğitim Bakanlığı'na bağlı resmi ve özel ilköğretim kurumlarında öğrenci başarısı, ilköğretim kurumları yönetmeliği dördüncü kısımda yer alan “Öğrenci Başarısının Değerlendirilmesi” başlığı altındaki ölçütler doğrultusunda belirlenmektedir (Milli Eğitim Bakanlığı [MEB], 2003). Bu doğrultuda ölçüt dayanaklı değerlendirme kullanılmakta ve 45 puan başarı ölçütü olarak dikkate alınmaktadır. Benzer şekilde ortaöğretim kurumlarında öğrenci başarısıyla ilgili karar alırken ölçüt dayanaklı değerlendirme kullanılmakta ve başarı ölçütü olarak yine 45 puan dikkate alınmaktadır. Bu duruma ilişkin detaylar ise ortaöğretim kurumları sınıf geçme ve sınav yönetmeliği çerçevesinde belirlenmektedir (MEB, 2004).

Yükseköğretim kurumlarında ise öğrenci başarısıyla ilgili kararlar alırken üniversitelerin eğitim-öğretim ve sınav yönetmeliklerinde yer alan kurallar dikkate alınmaktadır. Söz konusu yönetmeliklerde üniversite tarafından belirlenen kurallar doğrultusunda öğrenci başarısına ilişkin kullanılacak değerlendirme yöntemi, “Başarının Değerlendirilmesi” başlığı altında açıklanmaktadır. Ayrıca bir öğrencinin başarılı sayılabilmesi için gerekli olan alt limit puanı da söz konusu başlık altında belirtilmektedir. Yönetmeliklerde yer alan değerlendirme yöntemleri ve başarılı sayılmak için gerekli olan minimum puanlar incelendiğinde ise üniversiteler arasında farklılıklar olduğu dikkat çekmektedir.

Yükseköğretim kurumlarının ön lisans ve lisans eğitim-öğretim ve sınav yönetmeliklerinde yer alan başarı notlarının hesaplanması incelendiğinde farklı yöntemlerin kullanıldığı görülmektedir (Toprakçı ve diğerleri, 2007). Bazı üniversiteler norm dayanaklı değerlendirme kullanırken bazıları ise ölçüt dayanaklı değerlendirme kullanmaktadır. İki farklı değerlendirme yönteminin de kullanıldığı bazı üniversitelerde ise kullanılan yöntem üniversite birimleri arasında bile değişiklik gösterebilmektedir.

Yükseköğretim kurumlarında başarılı sayılmak için gerekli olan minimum puanlar incelendiğinde ise yine farklılıklar öne çıkmaktadır (Toprakçı ve diğerleri, 2007). Bazı üniversitelerde bir dersten başarılı sayılmak için 60 ve üzeri not almak gerekirken diğer üniversitelerde 50'den düşük olan puanlar başarısız sayılmakta ve ders tekrarını gerektirmektedir. Bunun yanında norm ve ölçüt dayanaklı değerlendirmelerin birlikte yapıldığı bazı üniversitelerde başarılı sayılmak için gerekli olan minimum puan, üniversite birimleri arasında değişiklik göstermektedir. Bu doğrultuda bir dersi başarıyla tamamlamak için bağlı

değerlendirme yapılan derslerde 60 ve üzeri, ölçüt dayanaklı değerlendirme yapılan derslerde 50 ve üzeri puan almak gerekmektedir.

Yükseköğretim kurumlarında kullanılan değerlendirmelerde ve başarı ölçütlerinde öne çıkan farklılıklar başarıya ilişkin alınan kararlara yansiyabilmektedir. Bir öğrencinin norm dayanaklı değerlendirmeden alacağı 55 puan, bazı üniversitelerde başarısız sayılmakta ve ders tekrarı kararına neden olmaktadır. Söz konusu puan, bağıl değerlendirmenin yapıldığı başka üniversitelerde ise şartlı başarılı sayılmakta ve öğrencinin diğer derslerdeki başarısına göre ders tekrarı kararı verilmemektedir.

Eğitim sisteminde değerlendirme yöntemlerinde öne çıkan farklılıklar ülkemizdeki diğer devlet kurumlarında ve özel sektörde de gözlemlenmektedir. Bireylerin seçiminde ya da verilen eğitimlerin etkinliğinin değerlendirilmesinde kurumlar norm veya ölçüt dayanaklı değerlendirmeler yapmakta ve hedefler benzerlik göstermesine rağmen kullanılan değerlendirme yöntemi kurumdan kuruma değişmektedir. Söz konusu değişkenlik nedeniyle değerlendirme öznel bir süreç haline gelmektedir (Özgüven, 2011). Bu doğrultuda ülkemizdeki kurumların değerlendirme yöntemlerinden hangisinin hangi durumlar için uygun olduğu konusunda yeterince bilgiye sahip olmadıkları gözlemlenmektedir.

Değerlendirme yöntemlerinin avantajları ve dezavantajları hakkında yeterli bilgiye sahip olmamak bir takım sorunlara neden olmaktadır. Kurumlar bireylerin seçiminde yanlış kararlar alabilmekte ve bu nedenle ekonomik zararlara uğrayabilmektedir. Verilen eğitimlerin hedeflere ne ölçüde ulaştığını değerlendirirken de yanlış kararlar alınabilmektedir. Alınan yanlış kararlar ise bir sonraki hedefe ulaşma noktasında sorunlara neden olabilmektedir.

Değerlendirmeye tabi tutulan birey açısından daha adil ve değerlendirmeyi yapan kurum açısından daha etkin kararlar alabilmek için, değerlendirme yöntemlerine göre kararların değişip değişmediğini araştırmak gerekmektedir. Alınan kararların norm veya ölçüt dayanaklı değerlendirme yöntemlerine göre değiştiği sonucuna ulaşılması durumunda ise hangi değerlendirme yöntemini kullanmanın daha geçerli olduğunu incelemek gerekmektedir. Bu araştırma ile aynı kişilerin aynı kapsam alanındaki başarı düzeylerine ilişkin norm ve ölçüt dayanaklı değerlendirme yöntemleriyle alınan kararlar karşılaştırılarak uygulamalardaki ihtiyaca cevap vermesi hedeflenmektedir.

1. 2. PROBLEM CÜMLESİ

I) Norm ve ölçüt dayanaklı değerlendirmeler sonucunda geçen öğrenci yüzdeleri arasında manidar fark var mıdır?

1. Angoff standart belirleme yöntemine ve 60 T puanı ölçütüne göre geçen öğrenci yüzdeleri arasında manidar fark var mıdır?
2. Angoff standart belirleme yöntemine ve 50 T puanı ölçütüne göre geçen öğrenci yüzdeleri arasında manidar fark var mıdır?
3. Nedelsky standart belirleme yöntemine ve 60 T puanı ölçütüne göre geçen öğrenci yüzdeleri arasında manidar fark var mıdır?
4. Nedelsky standart belirleme yöntemine ve 50 T puanı ölçütüne göre geçen öğrenci yüzdeleri arasında manidar fark var mıdır?
5. 50 ve 60 T puanlarına göre geçen öğrenci yüzdeleri arasında manidar fark var mıdır?

II) Norm ve ölçüt dayanaklı değerlendirmelerin uygunluk geçerlikleri nasıldır?

1. 60 T puanı ölçütüne göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyum nasıldır?
2. 50 T puanı ölçütüne göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyum nasıldır?
3. Angoff standart belirleme yöntemine göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyum nasıldır?
4. Nedelsky standart belirleme yöntemine göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyum nasıldır?

1. 3. SINIRLILIKLAR

Bu araştırma,

1. Başarının belirlenmesinde kullanılan 50 ve 60 bağıl ölçütleri
2. Standart belirleme yöntemlerinden test merkezli yöntemler olan Angoff ve Nedelsky Yöntemleri
3. Yöntemlerin uygulandığı lisans düzeyinde ölçme ve değerlendirme dersi
4. 2012 – 2013 eğitim ve öğretim döneminde verilerin toplandığı Kayseri Erciyes Üniversitesi, rehberlik ve psikolojik danışmanlık bölümü, 2. sınıf öğrencileri
5. Geçerlik türlerinden uygunluk geçerliği ile sınırlandırılmıştır.

1. 4. SAYILTILAR

1. Angoff ve Nedelsky Yöntemleri uygulanırken uzmanların test maddeleri ile ilgili kararlarını verirken yalnızca minimum yeterlikteki bireylerin öğrenme düzeylerini dikkate aldıkları varsayılmıştır.

1. 5. ARAŞTIRMANIN AMACI VE ÖNEMİ

Bu araştırma, aynı kişilerin aynı alandaki başarı düzeylerine ilişkin kararların kullanılan değerlendirme yöntemlerine göre anlamlı düzeyde değişip değişmediğini incelemek amacıyla planlanmıştır. Araştırmanın bir başka amacı ise kullanılan değerlendirme yöntemleri doğrultusunda alınan kararlarda farklılığa ulaşılması durumunda hangi yöntemin ya da yöntemlerin daha geçerli olduğunu belirlemektir.

Bu araştırma, uygulamada kullanılan değerlendirme yöntemlerinin farklılığına ve söz konusu farklılıkların bireylerle ilgili kararlara yansıyor yansımadığına dikkat çekmek adına önemlidir. Ayrıca hangi değerlendirme yöntemiyle karar almanın değerlendirmeye tabi tutulan birey açısından daha adil olacağını ve değerlendirmeyi yapan kurum açısından daha etkin sonuçlar almasını sağlayacağını göstermesi bakımından önemli olduğu düşünülmektedir. Değerlendirme yöntemlerinin nasıl ve hangi durumlarda uygulanacağına ilişkin alandaki bilgi ihtiyacını karşılaması açısından önemli bir araştırmadır. Bunun yanında standart belirleme yöntemleriyle yapılan değerlendirmeler norm dayanaklı değerlendirmelere oranla daha fazla zaman ve iş gücü gerektirdiği için çalışmada iki yöntem açısından bulunabilecek olası benzerlikler bu zaman ve iş gücü kaybını önlemesi açısından önemlidir.

1. 6. İLGİLİ ARAŞTIRMALAR

MacTurk ve Neisworth (1978) engeli olmayan ve engelli çocukların iletişim, entellektüel, sosyal ve motor gelişimlerini değerlendiren norm dayanaklı ölçme aracı ile ölçüt dayanaklı ölçme aracını karşılaştıran bir çalışma yapmıştır. Çalışmaya söz konusu gelişim alanlarından en az ikisinde bozukluğa sahip 7 engelli çocuk ile 10 engelsiz çocuk olmak üzere yaşları 19 ile 51 ay arasında değişen 17 çocuk katılmıştır. Norm dayanaklı test ile ölçüt dayanaklı test 3 ay arayla uygulanarak sonuçlar arasındaki ilişki Spearman sıra farkı korelasyon katsayısı ile incelenmiştir. Tüm grubun verileri ve sadece engelsiz çocukların verileri dikkate alındığında norm dayanaklı test ile ölçüt dayanaklı testin puanları arasındaki ilişki düşük bulunmuştur. Ancak sadece engelli çocukların verileri dikkate alındığında iki testin alt testleri arasında 0,93 ve 0,99 arasında değişen yüksek ilişkiler olduğu sonucuna ulaşılmıştır.

Oescher, Kirby ve Paradise (1992) norm dayanaklı ve ölçüt dayanaklı başarı testlerini karşılaştırmak adına okul çağındaki çocuklara ilişkin bir araştırma yapmıştır. Üçüncü ve onuncu sınıfa giden 29.000 öğrenci ile yapılan çalışmada her bir öğrenciye okuma, dil ve matematik alanlarındaki başarıyı ölçen bir norm dayanaklı test ile bir ölçüt dayanaklı test uygulanmıştır. İki farklı değerlendirme yöntemiyle elde edilen sonuçlar arasındaki ilişki Pearson momentler çarpımı korelasyon katsayısı ile incelenmiştir. Sonuç olarak iki farklı değerlendirme yöntemine ilişkin alt testlerin puanları arasında 0,70 ve 0,80 arasında değişen yüksek ilişkiler olduğu bulunmuştur.

Behuniak ve Tucker (1992) diğer araştırmacılardan farklı olarak norm dayanaklı değerlendirmelerle ölçüt dayanaklı değerlendirmeler arasında psikometrik bir ilişki bulmaya hedeflemiştir. Bu amaçla ölçüt dayanaklı bir test üzerinden test eşitleme yöntemleriyle norm dayanaklı test sonuçları kestirilmeye çalışılmıştır. Çalışma doğrultusunda matematik, okuma ve dil alanlarını değerlendiren bir ölçüt dayanaklı test ile bir norm dayanaklı test; dördüncü, altıncı ve sekizinci sınıflara giden öğrencilere uygulanmıştır. Çalışmaya her bir sınıf düzeyinden 35.000 öğrenci katılmıştır. Sonuç olarak test eşitleme yöntemleriyle ölçüt dayanaklı test üzerinden elde edilen puanların doğruluğunun yüksek olduğu bulunmuştur. Ayrıca eşitleme yoluyla elde edilen puanlar norm dayanaklı gerçek puanlar ile değiştirildiğinde elde edilen eşitleme hatası, norm dayanaklı paralel forma göre elde edilen hata ile benzerlik göstermiştir.

Visintainer (2002), öğrencilerin akademik başarılarını ölçmek için geliştirilmiş norm dayanaklı test ile ölçüt dayanaklı test arasındaki ilişkiyi araştıran bir çalışma yapmıştır. Araştırmaya katılan 1129 dördüncü ve beşinci sınıf öğrencisine ilk olarak norm dayanaklı test uygulanmıştır. Bir yıl sonra öğrencilere ölçüt dayanaklı test uygulanarak norm dayanaklı test ile ölçüt dayanaklı test arasındaki ilişki incelenmiştir. İki farklı değerlendirmeden elde edilen alt testlerin

puanları arasında Pearson momentler çarpımı korelasyon katsayısına göre 0,43 ve 0,69 arasında değişen orta seviyede ilişki olduğu bulunmuştur. Ayrıca yapılan çoklu regresyon analizi sonucuna göre norm dayanaklı testin ölçüt dayanaklı testi yordadığı bulunmuştur. Bu doğrultuda norm dayanaklı testin ölçüt dayanaklı teste ait matematik alt testini % 51,3 açıkladığı sonucuna ulaşılmıştır.

Pester (2003), dil gelişimini değerlendiren norm dayanaklı testlerle ölçüt dayanaklı testleri karşılaştırmak istemiştir. Bu doğrultuda dil gelişimi geciken çocuklara bir eğitim verilmiştir ve çocukların gelişimini değerlendirmek için ön ve son testler uygulanmıştır. Çalışmaya dil gelişiminde sorun yaşayan ve yaşları 25-54 ay arasında değişen 18 çocuk katılmıştır. Beş aylık eğitimin öncesinde ve sonrasında anlama ve kendini ifade etme alt testleri olan bir ölçüt dayanaklı test ile bir norm dayanaklı test uygulanmıştır. Bağımlı gruplar t-testi sonuçları iki farklı test için de manidar bulunmuş ve iki farklı test açısından da çocuklarda gelişim olduğu sonucuna ulaşılmıştır. Alt testler açısından ise her iki değerlendirme için de ifade etme alt testinde gelişim olduğu; ancak anlama alt testinde gelişim olmadığı bulunmuştur.

Çetin (2006), öğrencilerin geçti/kaldı durumlarının farklı standart belirleme yöntemlerine göre değişiklik gösterip göstermediğini araştırmıştır. Bu doğrultuda 25 maddelik Türkçe ve Matematik testleri 129 ilköğretim 7. sınıf öğrencisine uygulanmıştır. Z testi sonuçları doğrultusunda Angoff, Nedelsky ve Yes/No standart belirleme yöntemlerine göre başarılı olan öğrenci yüzdeleri her iki ders için de farklılık göstermiştir. Nedelsky Yöntemi'ne göre başarılı olan öğrenci yüzdelerinin diğer yöntemlere göre daha yüksek olduğu bulunmuştur. Derslerin sayısal ya da sözel içerikli olmasına göre başarılı olan öğrenci yüzdeleri arasında ise önemli farklılıklar olduğu sonucuna ulaşılmıştır.

Mohr (2006); dil, okuma ve matematik alt testleri olan ölçüt dayanaklı başarı testi ile norm dayanaklı başarı testi arasındaki ilişkiyi incelemiştir. Bu doğrultuda ilk olarak norm dayanaklı test olmak üzere iki test 4.000 dördüncü sınıf öğrencisine uygulanmıştır. Ayrıca ölçüt dayanaklı testin alt testlerindeki başarı dört grupta sınıflanmış ve bağımsız değişken olarak dikkate alınan norm dayanaklı testin diskriminant analizi sonucunda ölçüt dayanaklı testi ne kadar açıkladığı incelenmiştir. Sonuç olarak norm dayanaklı test ile ölçüt dayanaklı testin alt testlerine ilişkin puanlar arasında orta düzeyde ilişki olduğu bulunmuştur. Bunun yanında diskriminant analizine göre ölçüt dayanaklı teste ait özellikle uç noktadaki başarı sınıflarını, norm dayanaklı testin yüksek oranlarda açıkladığı sonucuna ulaşılmıştır.

Nartgün (2007) araştırmasında, üç farklı derse ilişkin notlar üzerinden mutlak ve bağıl değerlendirme yöntemlerinin öğrencilerin başarı düzeyleri sayılan notlarda farklılık yaratıp yaratmadığını incelemiştir. X, Y ve Z dersleri için sırasıyla 63, 28 ve 8 öğrenciye ait hipotetik

puanlar türetilmiştir. Söz konusu puanlar, mutlak değerlendirme kullanan bir üniversitenin değerlendirme yönetmeliği ve bağıl değerlendirme kullanan bir üniversitenin değerlendirme yönetmeliği doğrultusunda başarı düzeylerine dönüştürülmüştür. Dönüştürülen puanların iki farklı değerlendirme yönteminde hangi başarı düzeyine denk geldiği karşılaştırılmıştır. Özellikle notların düşük olduğu derste, mutlak değerlendirme yöntemiyle elde edilen başarı düzeyleri düşük olan öğrencilerin, bağıl değerlendirmede çok daha yüksek başarı düzeylerine sahip oldukları bulunmuştur. Çalışmada dikkate alınan derslerde iki farklı değerlendirme yöntemine ilişkin başarı düzeylerinde farklılıklara ulaşılmıştır.

Jacobson (2008), öğretmenler tarafından geliştirilen norm dayanaklı test ile ölçüt dayanaklı testin okuma ve matematik alanlarındaki öğrenci performansını sınıflama başarılarını karşılaştırmıştır. Bu doğrultuda dördüncü, sekizinci ve on ikinci sınıf seviyeleri dikkate alınarak değerlendirme yöntemlerinin başarısına ilişkin 112 sınıf öğretmeninin görüşüne başvurulmuştur. Sonuç olarak her iki değerlendirme yönteminin de öğrencileri yeterli ve yetersiz olmak üzere iki kategoride başarılı bir şekilde sınıfladığı bulunmuştur. Ancak sınıflama kategorisi dörde çıkarıldığında her iki değerlendirme yönteminin de en iyi kategori olan uzman seviyesini yeterince başarılı bir şekilde sınıflayamadığı bulunmuştur. Matematik alanı açısından 2 seviyeli performans sınıflaması dikkate alındığında norm dayanaklı değerlendirmenin, 4 performans sınıflaması dikkate alındığında ölçüt dayanaklı değerlendirmenin daha başarılı olduğu sonucuna ulaşılmıştır. Okuma alanı açısından ise norm dayanaklı değerlendirmenin öğrencileri daha başarılı sınıfladığı bulunmuştur.

Taşdelen (2009), Nedelsky ve Angoff standart belirleme yöntemleriyle elde edilen kesme puanlarını genellenebilirlik kuramını kullanarak karşılaştırmıştır. Bu doğrultuda 415 öğrenciye altıncı sınıf fen ve teknoloji dersine ilişkin 16 maddelik test uygulanmıştır. 40 fen ve teknoloji uzmanının görüşleri doğrultusunda iki farklı standart belirleme yöntemine ilişkin başarı yüzdelerinin farklılaştığı sonucuna ulaşılmıştır. Nedelsky Yöntemi'ne ilişkin puanlayıcılar arasında kesme puanı belirlemede uyum olmadığı görülmüş; ancak Angoff Yöntemi'nde puanlayıcılar arasında bir tutarlılığın olduğu bulunmuştur. Angoff ve Nedelsky yöntemleri ile standart belirleme çalışırken 10 puanlayıcının yeterli olduğu sonucuna varılmıştır.

Çetin (2011), İşaretleme ve Angoff standart belirleme yöntemlerini bir üniversitenin mevcut değerlendirme yöntemiyle karşılaştırmasını içeren bir çalışma yapmıştır. İngilizce seviyesini ölçen 55 soruluk dilbilgisi ve sözcük bilgisi maddelerine ilişkin 17 dil uzmanının görüşlerine başvurmuştur. Z testi sonuçları doğrultusunda üniversitenin mevcut değerlendirme sistemine göre başarılı olan öğrenci yüzdeleri ile standart belirleme yöntemlerine göre başarılı olan öğrenci yüzdelerinin farklılaştığı sonucuna ulaşılmıştır. İşaretleme ve Angoff standart belirleme

yöntemlerine göre başarılı olan öğrenci yüzdeleri arasında ise önemli bir fark olmadığı saptanmıştır.

Gündeğer (2012); Angoff, Yes/No ve Ebel standart belirleme yöntemleri ile elde edilen kararları ve kesme puanları karşılaştıran bir araştırma yapmıştır. Bu doğrultuda matematik dersine ilişkin başarı testi dördüncü sınıfa giden 489 öğrenciye uygulanmıştır. Uzman görüşleriyle hesaplanan üç yönteme ilişkin kesme puanlarına göre başarılı sayılan öğrenci yüzdelerinin z testi doğrultusunda farklılık gösterdiği bulunmuştur. Öğrencilerin başarılı ve başarısız olarak sınıflanma durumları açısından ise yöntemler arasında uyum olduğu sonucuna ulaşılmıştır.

Gündeğer ve Doğan (2012); Angoff, Yes/No ve Ebel standart belirleme yöntemleriyle elde edilen üç farklı geçme puanı ile Milli Eğitim Bakanlığı tarafından ilköğretim okullarında 45 olarak tanımlanan geçme puanına göre başarılı sayılan öğrenci yüzdelerini karşılaştıran bir çalışma yapmıştır. Ayrıca çalışmada, geçme puanlarına göre öğrencilerin başarılı-başarısız sayılma durumları bakımından uyum olup olmadığı da incelenmiştir. Bu doğrultuda 20 maddelik matematik başarı testi, 489 ilköğretim dördüncü sınıf öğrencisine uygulanmıştır. Yöntemlere ilişkin başarı yüzdeleri 45 geçme puanına göre başarı yüzdesiyle ikili olarak karşılaştırıldığında, z testi sonucuna göre manidar farklılıklar bulunmuştur. Ayrıca Cohen'in Kappa istatistiğine göre öğrencilerin başarılı-başarısız sayılma durumları bakımından geçme puanları arasında anlamlı bir uyum olduğu sonucuna ulaşılmıştır.

Farklı testler üzerinden norm ve ölçüt dayanaklı değerlendirmeleri karşılaştıran araştırmalar incelendiğinde çoğunun iki farklı değerlendirme arasında ilişki bulunduğu görülmektedir. Standart belirleme yöntemleri açısından ise farklı yöntemlere göre alınan başarı kararlarına ilişkin yüzdelerin, araştırmaların çoğunda farklılaştığı görülmektedir. Bu araştırma kapsamında da değerlendirilen özellikle Angoff ve Nedelsky yöntemlerinin birçok araştırmaya konu olduğu dikkat çekmektedir.

Yapılan yurt dışı araştırmalar incelendiğinde ise norm ve ölçüt dayanaklı değerlendirme yöntemlerinin farklı testler üzerinden karşılaştırılmasına yönelik çalışmalar olduğu görülmektedir. Ayrıca söz konusu çalışmalarda alınan kararlar arasındaki farklılıkların da araştırıldığı görülmektedir. Ancak ölçüt dayanaklı değerlendirmelerde kullanılan standart belirleme yöntemleri ile norm dayanaklı değerlendirmelerin karşılaştırılmadığı dikkat çekmektedir. Bunun yanında aynı test üzerinden iki değerlendirme yönteminin karşılaştırıldığı çalışmalara ulaşılamamıştır. Bu araştırmada ise aynı test üzerinden norm ve ölçüt dayanaklı değerlendirmeler karşılaştırılmıştır.

Yurt içinde yapılan araştırmalar incelendiğinde ise standart belirleme yöntemleri kendi içerisinde karşılaştırılmış ya da bir başka ölçüt dayanaklı değerlendirmeyle karşılaştırılmıştır.

Ancak norm dayanaklı deęerlendirmelerle karřılařtırılmadıęı grlmektedir. Bu alıřma konusuna benzer řekilde sadece Nartgn (2007) arařtırmasında iki farklı deęerlendirme yntemini karřılařtırmıřtır. Ancak Nartgn'n alıřmasından farklı olarak bu alıřmada lt dayanaklı deęerlendirme iin dikkate alınan yntem standart belirleme yntemleri olmuřtur.

zellikle norm dayanaklı deęerlendirmelerin uygulamada sıklıca bařvurulan bir yntem olması ve standart belirleme yntemlerinin de son yıllarda birok arařtırmaya konu olması nedeniyle standart belirleme yntemleri kullanılarak elde edilecek kesme puanına gre karar vermenin norm dayanaklı deęerlendirmelerle verilen kararlarla karřılařtırılması nem kazanmaktadır. Bu arařtırmada alandaki alıřmalardan farklı olarak aynı test zerinden lt dayanaklı deęerlendirme olan standart belirleme yntemleriyle norm dayanaklı deęerlendirmeler karřılařtırılmıřtır.

Arařtırmaların biroęunda iki farklı deęerlendirme ynteminin sadece karřılařtırılması ise dikkat eken bir bařka noktadır. Norm ve lt dayanaklı deęerlendirmelerin farklılıkları ve benzerlikleri arařtırılmasına raęmen hangi yntemin daha geerli olduęu sorusu yanıtlanmamıřtır. Sadece Jacobson (2008) norm ve lt dayanaklı deęerlendirmelerin doęru sınıflama bařarılarını karřılařtırdıęı grlmektedir. Bu nedenle bu arařtırmada, iki farklı deęerlendirme ynteminde bulunacak olası farklılıklarda hangi yntemin daha geerli olduęu sorusuna da cevap aranılmıřtır.

BÖLÜM II

YÖNTEM

Bu bölümde araştırmanın türü, verilerin elde edildiği çalışma grubu, verilerin toplanması ve verilerin analizinde kullanılan istatistiksel yöntemler açıklanmıştır.

2. 1. ARAŞTIRMANIN TÜRÜ

Norm dayanaklı değerlendirmeler ile ölçüt dayanaklı değerlendirmelerin geçme kararı üzerine etkisinin araştırıldığı bu çalışmada değişkenlerin birlikte değişimi tespit edilmeye çalışıldığı için araştırma türü, ilişkisel türden araştırma modelidir. Ayrıca söz konusu değerlendirmelerin ilgili dersten yıl sonu geçme ve kalma kararlarıyla uyumunun araştırılması açısından da ilişkisel türden bir araştırma modelidir. Değerlendirme yöntemlerinin özelliklerinin belirlenmesi açısından ise betimsel bir çalışmadır.

2. 2. ÇALIŞMA GRUBU

Araştırmada elde edilen sonuçlar herhangi bir şekilde evrene genellenmeyeceği için evren örneklem ilişkisine gidilmemiştir. Bu doğrultuda amaca uygun çalışma grubu seçilmiştir. Çalışma grubunu Kayseri Erciyes Üniversitesi rehberlik ve psikolojik danışmanlık bölümü 2. sınıfında okuyan öğrenciler oluşturmuştur. Söz konusu çalışma grubunun seçilmesinde öğrencilerin ölçme ve değerlendirme dersi almış olmaları etkili olmuştur. Ayrıca standart belirleme yöntemlerinde uzman görüşü olarak ölçme ve değerlendirme alanında en az yüksek lisans eğitimi almış uzmanların görüşlerine başvurulmuştur.

Tablo 2: Araştırmaya Katılan Öğrenci ve Uzman Sayıları

	Kişi Sayısı
Öğrenci	107
Uzman (Angoff)	11
Uzman (Nedelsky)	10

Tablo 2'ye bakıldığında araştırmaya rehberlik ve psikolojik danışmanlık 2. sınıf öğrencilerinden 107 kişi katıldığı görülmektedir. Uygulamanın yapıldığı gün bazı öğrencilerin devamsızlık yapması nedeniyle öğrencilerin tamamına ulaşılamamıştır. Uzman görüşü olarak ise Angoff

Yöntemi için 11 ve Nedelsky Yöntemi için 10 alan uzmanına ulaşılmıştır. 7 uzmandan hem Angoff hem de Nedelsky yöntemleri için görüş alınmıştır. 4 uzman sadece Angoff Yöntemi ve 3 uzman sadece Nedelsky Yöntemi için görüş belirtmiştir. Uzman sayısı belirlenirken, Taşdelen'in (2009) Angoff ve Nedelsky yöntemleri ile standart belirleme çalışırken 10 puanlayıcının yeterli olduğunu gösteren araştırma sonucu dikkate alınmıştır.

2. 3. VERİLERİN TOPLANMASI VE VERİ TOPLAMA ARACI

Araştırma kapsamında kullanılan veriler üç farklı süreçte toplanmıştır. İlk süreci öğrencilerden elde edilen veriler oluşturmaktadır. Bu amaçla ölçme ve değerlendirmeye ilgili hazırlanan 20 soruluk bir başarı testi kullanılmıştır. İkinci süreçte ise Angoff ve Nedelsky standart belirleme yöntemleri doğrultusunda testin kesme puanını hesaplamak için uzman görüşlerine başvurulmuştur. Uzman görüşleri doğrultusunda başarı testinin kesme puanları elde edilmiştir. Son süreçte ise uygunluk geçerliğini incelemek için öğrencilerin ilgili dersten yıl sonunda aldıkları notlar kullanılmıştır.

Çalışmanın ilk süreci, ölçme ve değerlendirmeye ilgili bir başarı testi geliştirme ve başarı testini çalışma grubuna uygulama adımlarından oluşmaktadır. Literatürde yer alan çalışmaların aksine bu çalışmada iki yerine bir test üzerinden yöntemler karşılaştırılmıştır. Madde güçlükleri, madde ayırt edicilikleri, maddelerde gözlemlenen davranışlar, testin kapsamı gibi madde ve test özellikleri açısından eşdeğer iki test geliştirme zorluğu nedeniyle bu çalışmada sadece bir test kullanılmıştır. Bu sayede değerlendirme yöntemleri dışında başka faktörlerin çalışma sonuçlarında etkili olmasının önüne geçilmek istenmiştir.

Çalışmada kullanılan testin kapsamı belirlenirken Aktan ve Çepni'nin (2010) çalışmasından faydalanılmıştır. Söz konusu çalışmada uzman görüşleri çerçevesinde ölçme ve değerlendirme kapsamı ve gereklilikleri araştırılmıştır. Testin kapsamı belirlenirken kullanılan bir diğer ölçüt ise Kayseri Erciyes Üniversitesi ölçme ve değerlendirme ders tanımı olmuştur. Belirlenen konu başlıkları doğrultusunda Kamu Personeli Seçme Sınavı'nda (KPSS) çıkan sorulardan faydalanılmıştır. Soruların konu başlıklarına göre dağılımında ise KPSS'de çıkan soruların konulara göre ağırlıkları dikkate alınmıştır. Seçilen sorular ölçme ve değerlendirme programında doktora yapan 2 uzman tarafından incelenerek ön uygulama için 36 soruluk test formu oluşturulmuştur.

Oluşturulan test formunun madde analizleri için Kayseri Erciyes Üniversitesi'nde 30 kişilik bir gruba ön uygulaması yapılmıştır. Ön uygulamaya rehberlik ve psikolojik danışmanlık 2. sınıf öğrencileri katılmıştır. Ön uygulamadan elde edilen veriler doğrultusunda ITEMAN programı

kullanılarak madde analizleri yapılmıştır. Madde analizlerinde maddelerin güçlük ve ayırt edicilik indeksleri kullanılmıştır. Madde ayırt edicilik indeksinin belirlenmesinde ise çift serili korelasyon katsayısı dikkate alınmıştır. Madde ayırt edicilik gücünde ölçüt olarak 0,30 kullanılmıştır. Madde güçlük indeksi açısından ise 0,50'ye yakın olan orta güçlükteki maddelere ağırlık verilmeye çalışılmıştır. Testin kapsam alanı, maddelerin güçlük ve ayırt edicilik indeksleri doğrultusunda gerekli düzeltmeler yapıldıktan sonra her biri beş seçenekli 20 soruluk başarı testi (Ek 1) oluşturulmuştur. Başarı testinin konulara göre dağılımı Tablo 3'de sunulmuştur.

Tablo 3: Veri Toplama Aracının Konu Başlıkları ve Soru Sayılarına Göre Dağılımı

Konu Başlığı	Alt Başlık	Soru Sayısı	Soru Numarası
Ölçme ve Değerlendirmede Temel Kavramlar	Eğitimde Ölçme ve Değerlendirmenin Önemi	1	9
	Ölçüt ve Ölçüt Çeşitleri	1	10
	Ölçmede Sıfır	1	1
	Değişken ve Çeşitleri	1	11
Ölçek ve Ölçek Türleri	Ölçek Türleri Nelerdir	1	13
Değerlendirme ve Değerlendirme Türleri	Değerlendirme Türleri	1	2
	Amacına Göre Yapılan Değerlendirme	1	15
	Ölçüte Göre Yapılan Değerlendirme	1	14
Ölçme Aracının Özellikleri	Ölçmede Hata	1	17
	Geçerlik	1	3
	Güvenirlilik	1	16
Test Hazırlama Süreci ve Ölçme Araçları	Test Hazırlama Süreçleri	1	5
	Ölçme Araçları	2	4,6
Ölçmede İstatistiksel İşlemler	Madde İstatistikleri	3	18,19,20
	Test İstatistikleri	3	7,8,12
Toplam		20	

Oluşturulan beş seçenekli 20 soruluk başarı testinin nihai uygulaması Kayseri Erciyes Üniversitesi'nde yapılmıştır. Uygulamaya söz konusu üniversitenin rehberlik ve psikolojik danışmanlık bölümü 2. sınıf öğrencilerinden 107 kişi katılmıştır. Uygulama 2013 yılının Mart ayında gerçekleşmiştir. Uygulamadan elde edilen veriler doğrultusunda madde ve test istatistikleri incelenmiştir. Bu doğrultuda başarı testinin ITEMAN programında madde analizleri (Ek 3) yapılmıştır. Yapılan analizler sonucunda test maddelerinin ayırt edicilik gücü

indekslerinin 0,33 ile 0,80 arasında deđiřtiđi bulunmuřtur. Maddelerin gclk indekslerinin ise 0,11 ile 0,85 arasında deđiřtiđi sonucuna ulařılmıřtır.

Tablo 4: Nihai Uygulama Sonularına Gre Testin zellikleri

k	n	S _x	\bar{x}	\bar{p}	arpıklık	Basıklık	K-Smirnov Z	KR-20
20	107	3,39	11,08	0,55	-0,06	-0,26	0,76*	0,71

* $p > 0,05$

Test istatistiklerine iliřkin bilgiler Tablo 4’de verilmiřtir. Buna gre lme ve deđerlendirme bařarı testinin KR-20 ile hesaplanan i tutarlık gvenirliđi 0,71 bulunmuřtur. Testin ortalama gclđe ($\bar{p}=0,55$) sahip olduđu grlmektedir. arpıklık ve basıklık katsayılarının sifıra yakın olması nedeniyle dađılımın normal dađılıma uygun olduđu grlmektedir. Testten elde edilen puanların dađılımını incelemek iin yapılan Kolmogorov-Smirnov testine gre đrencilerin bařarı testi puanlarının normal dađılıma sahip olduđu ($Z=0,76$; $p > 0,05$) bulunmuřtur. Sonu olarak istatistiksel zellikler ve uzman grřleri dođrultusunda veri toplama aracı oluřturulmuřtur.

Veri toplama srecinin ikinci ařaması, Angoff ve Nedelsky standart belirleme yntemleri dođrultusunda kesme puanları hesaplamak iin uzman grřlerine bařvurmayı kapsamaktadır. Buna gre standart belirleme yntemlerinin uygulanabilmesi iin ilk olarak minimum yeterlik dzeyi incelenmiřtir. Uzmanlarla bir araya gelme gclđu nedeniyle minimum yeterlik dzeyi; Amerikan Ulusal Eđitimde lme Kurulu (NCME), Amerikan đretmenler Federasyonu (AFT) ve Amerikan Ulusal Eđitim Birliđi’nin (NEA) belirlediđi lme ve deđerlendirme alanındaki yeterlik standartlarından (AFT, NCME ve NEA, 1990) faydalanılarak tanımlanmıřtır. Buna gre minimum yeterlik, “Eđitimle ilgili kararlar iin uygun lme aralarını seebilen, gerektiđinde geliřtirebilen, sonuları puanlayabilen ve yorumlayabilen (deđerlendirebilen), sonulardan faydalanabilen kiři” dikkate alındıđında yeterli ile yetersiz arasındaki birey olarak belirlenmiřtir.

Minimum yeterlik seviyesi tanımlandıktan sonra standart belirleme yntemlerinin uygulanmasına geilmiřtir. Uzmanlarla bir hafta arayla iki grup toplantısı yapılmıřtır. İlk olarak Angoff Yntemi uygulanmıř ve uzmanlardan belirlenen minimum yeterlik dzeyindeki kiřiyi dikkate alarak her bir maddenin lisans dzeyindeki đrenciler tarafından dođru cevaplanabilme olasılıđına karar vermeleri istenmiřtir. Uygulamadan bir hafta sonra ise Nedelsky Yntemi ile ilgili uzman grřlerine bařvurulmuřtur. Bu srete ise uzmanlardan minimum yeterlikteki

kişiyi dikkate alarak her bir maddede yer alan yanlış seçeneklerden kaçının lisans düzeyindeki öğrenciler tarafından elenebileceğini tahmin etmeleri istenmiştir.

Veri toplama sürecinin son aşamasında ise NDD ve ÖDD ile ilgili yöntemlerin uygunluk geçerliklerini karşılaştırmak için çalışmaya katılan öğrencilerin ders notlarına ulaşılmıştır. Bu doğrultuda Öğrenci İşleri Daire Başkanlığı'ndan öğrencilerin 2012-2013 eğitim yılı ölçme ve değerlendirme dersi dönem sonu notları alınmıştır.

2. 4. VERİLERİN ANALİZİ

Araştırmanın her bir alt probleminin değişkenlerinin belirtilmesi ve söz konusu problemlerin çözümlenmesinde kullanılan istatistiksel yöntemler alt problemlerin sıralamasına göre verilmiştir. Alt problemlere ilişkin değişkenler incelenmeden önce ilk olarak öğrencilerin T puanları ve standart belirleme yöntemleri doğrultusunda testin kesme puanları hesaplanmıştır.

Testin puanlaması sürecinde her bir sorunun test puanına katkısı 1 puan olarak dikkate alınmıştır. Norma bağlı puanları hesaplamak için ilk olarak öğrencilerin cevap anahtarları doğrultusunda sorulara doğru yanıtları “1” ve yanlış yanıtları “0” olarak kodlanmıştır. Bu doğrultuda her bir öğrencinin testten elde ettikleri toplam doğru sayılarına ulaşılmıştır. Toplam doğru sayıları ise takip eden formül yardımıyla T puanına dönüştürülmüştür (Sönmez ve Alacapınar, 2011).

$$T_i = \left(\frac{X_i - \bar{x}}{S_x} \right) * 10 + 50$$

T_i : Ortalaması 50 ve Standart Sapması 10 olan Standart Puan

X_i : Bireyin testten elde ettiği toplam doğru sayısı

\bar{x} : Testin aritmetik ortalaması

S_x : Testin standart sapması

T puanları, öğrencilerin norm dayanaklı değerlendirme sonucu elde ettikleri puanları oluşturmuştur. Toplam doğru sayıları doğrultusunda elde edilen T puanları 60 ölçütüne göre başarılı (1) ve başarısız (0) olarak sınıflandırılmıştır. Benzer süreç 50 ölçütüne göre de uygulanmıştır. Sonuç olarak öğrencilerin norma bağlı puanlar doğrultusunda başarı durumlarına ulaşılmıştır.

Alt problemler analiz edilmeden önceki ikinci süreçte ölçüt dayanaklı değerlendirme doğrultusunda testin Angoff ve Nedelsky Yöntemleri'ne göre kesme puanlarını hesaplamak için uzman görüşlerine başvurulmuştur. Söz konusu uzmanların görüşleri doğrultusunda Angoff ve Nedelsky Yöntemleri'ne göre testin kesme puanları "Microsoft Excel" programı kullanılarak hesaplanmıştır. Her bir uzmanın maddeler için belirledikleri olasılıklar toplanarak söz konusu uzmana ilişkin kesme puanı belirlenmiştir. Her bir uzman için hesaplanan kesme puanlarının aritmetik ortalaması alınarak teste ilişkin kesme puanına ulaşılmıştır. Angoff ve Nedelsky Yöntemleri için aynı süreç uygulanmış ve teste ilişkin kesme puanları elde edilmiştir. Elde edilen kesme puanları kullanılarak öğrencilerin testten aldıkları ham puanlar başarılı (1) ve başarısız (0) olarak sınıflandırılmıştır. Bu sayede ölçüt dayanaklı değerlendirme doğrultusunda öğrencilerle ilgili başarı durumlarına ilişkin kararlara ulaşılmıştır.

2. 4. 1. Birinci Problem

2. 4. 1. 1. Birinci Alt Problem: Angoff standart belirleme yöntemine ve 60 T puanı ölçütüne göre geçen öğrenci yüzdeleri karşılaştırılmıştır. Bağımlı iki oran arasındaki fark test edildiği için kullanılan istatistiki yöntem z testi olmuştur. Önem düzeyi ise 0,05 alınmıştır.

Verilerin toplandığı aynı gruba ait iki farklı yüzde arasındaki manidar farkı ortaya koymak için z testi kullanılmaktadır (Calmorin ve Calmorin, 2007). Z istatistiği değişkenler arasındaki gözlenen oran farkının, iki oran arasındaki farkın standart hatasına bölünmesiyle hesaplanmaktadır (Jekel, 2007).

		Frekanslar				Oranlar		
		Yöntem II				Yöntem II		
		Kaldı	Geçti			Kaldı	Geçti	
Yöntem I	Geçti	A	B	A+B		a	b	p ₁
	Kaldı	C	D	C+D		c	d	q ₁
		A+C	B+D	N		q ₂	p ₂	1.00

Yukarıdaki tablolarda büyük harfler frekansları, küçük harfler ise bulunduğu kutucuğa denk gelen frekansın kişi sayısına (N) oranını ifade etmektedir. Bu doğrultuda z istatistiği, takip eden formüle göre hesaplanmaktadır (Aktaran: Gündeğer, 2012).

$$z = \frac{p_1 - p_2}{\sqrt{\frac{a+d}{N}}}$$

p_1 : Birinci yöntemle göre başarı oranı

p_2 : İkinci yöntemle göre başarı oranı

a: Birinci yöntemden başarılı olup ikinci yöntemde başarısız olan kişilerin oranı

d: Birinci yöntemden başarısız olup ikinci yöntemde başarılı olan kişilerin oranı

N: Kişi sayısı

2. 4. 1. 2. İkinci Alt Problem: Angoff standart belirleme yöntemine ve 50 T puanı ölçütüne göre geçen öğrenci yüzdeleri karşılaştırılmıştır. Bağımlı iki oran arasındaki fark test edildiği için kullanılan istatistikî yöntem z testi olmuştur. Önem düzeyi ise 0,05 alınmıştır.

2. 4. 1. 3. Üçüncü Alt Problem: Nedelsky standart belirleme yöntemine ve 60 T puanı ölçütüne göre geçen öğrenci yüzdeleri karşılaştırılmıştır. Bağımlı iki oran arasındaki fark test edildiği için kullanılan istatistikî yöntem z testi olmuştur. Önem düzeyi ise 0,05 alınmıştır.

2. 4. 1. 4. Dördüncü Alt Problem: Nedelsky standart belirleme yöntemine ve 50 T puanı ölçütüne göre geçen öğrenci yüzdeleri karşılaştırılmıştır. Bağımlı iki oran arasındaki fark test edildiği için kullanılan istatistikî yöntem z testi olmuştur. Önem düzeyi ise 0,05 alınmıştır.

2. 4. 1. 5. Beşinci Alt Problem: 50 ve 60 T puanlarına göre geçen öğrenci yüzdeleri karşılaştırılmıştır. Bağımlı iki oran arasındaki fark test edildiği için kullanılan istatistikî yöntem z testi olmuştur. Önem düzeyi ise 0,05 alınmıştır.

2. 4. 2. İkinci Problem

2. 4. 2. 1. Birinci Alt Problem: 60 T puanı ölçütüne göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyumu belirlerken Cohen'in Kappa testi yapılmıştır.

Kategorik iki değişken arasındaki uyumu ölçerken Cohen'in Kappa testi kullanılmaktadır. Cohen'in Kappa testi, iki değişkenin sadece değişim yönüne ilişkin benzerlik bilgisi vermemekte ve değişkenler arasındaki uyumun şans olasılığını dikkate almaktadır. Cohen'in Kappa katsayısının 0,4 ile 0,6 arasında olması uyumu göstermektedir. Katsayının 0,6 ile 0,75

arasında olması iyi uyumu ve 0,75'den yüksek olması mükemmel uyumu ifade etmektedir. Cohen'in Kappa katsayısı takip eden formüle göre hesaplanmaktadır (Clark-Carter, 2005).

$$K = \frac{\text{Sum } f_o - \text{Sum } f_e}{N - \text{Sum } f_e}$$

K : Cohen'in Kappa katsayısı

$\text{Sum } f_o$: Gözlemlenen uyumlu kararların toplam frekansı

$\text{Sum } f_e$: Beklenen uyum frekanslarının toplamı

N: Kişi sayısı

2. 4. 2. 2. İkinci Alt Problem: 50 T puanı ölçütüne göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyumu belirlerken Cohen'in Kappa testi yapılmıştır.

2. 4. 2. 3. Üçüncü Alt Problem: Angoff standart belirleme yöntemine göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyumu belirlerken Cohen'in Kappa testi yapılmıştır.

2. 4. 2. 4. Dördüncü Alt Problem: Nedelsky standart belirleme yöntemine göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyumu belirlerken Cohen'in Kappa testi yapılmıştır.

BÖLÜM III

BULGULAR VE YORUMLAR

Bu bölümde araştırma verileri, ikinci bölümde belirtilen istatistiksel yöntemler doğrultusunda alt problemlerin sırası dikkate alınarak analiz edilmiş ve analizler yorumlanmıştır.

3. 1. BİRİNCİ PROBLEME AİT BULGULAR VE YORUMLAR

Problem: Norm ve ölçüt dayanaklı değerlendirmeler sonucunda geçen öğrenci yüzdeleri arasında manidar fark var mıdır?

Birinci probleme ait alt problemlere ilişkin bulgulara ulaşmak için ilk olarak Angoff Yöntemi ve Nedelsky Yöntemi doğrultusunda kesme puanları hesaplanmıştır. Bu doğrultuda uzman görüşlerine başvurulmuştur. Angoff Yöntemi için uzman görüşleri Ek 4’de ve Nedelsky Yöntemi için uzman görüşleri Ek 5’de verilmiştir. Uzmanlar arasındaki uyumu incelemek için SPSS 18’den faydalanılarak Kendall’in uyum katsayısı (W) hesaplanmıştır. Uzmanlara ilişkin uyum katsayıları Tablo 5’de verilmiştir.

Tablo 5: Angoff ve Nedelsky Yöntemleri’ne Göre Uzmanlar Arası Uyum Katsayıları

	Uzman Sayısı	Kendall’s W
Angoff	11	0,45**
Nedelsky	10	0,44**

** p < 0,01

Kendall’in uyum katsayısı (W), puanlayıcıların verdikleri puanlar arasında ne ölçüde uyum olduğunu gösteren bir yöntemdir (Deliceoğlu, 2009). İki den fazla puanlayıcı olması durumunda kullanılan ve puanlayıcılar arası güvenilirliği belirleyen yöntem, parametrik olmayan bir tekniktir (Gündeğer, 2012). Angoff Yöntemi’ne katılan 11 uzmanın uyumu incelendiğinde Kendall’in W uyum katsayısının 0,01 hata düzeyinde manidar olduğu görülmektedir. Kendall’in W katsayısı doğrultusunda uzmanlar arasında uyum (W = 0,45) olduğu görülmektedir. Benzer şekilde Nedelsky Yöntemi’ne katılan 10 uzman arasında da Kendall’in W katsayısına göre 0,01 hata düzeyinde uyum olduğu (W = 0,44) sonucuna ulaşılmıştır. Uzmanların aynı ortamda görüşlerinin alınması, uzmanlar arası uyumu etkilemiş olabilir. Uzmanlar arasındaki uyum incelendikten sonra Angoff ve Nedelsky Yöntemleri için kesme puanları hesaplanmıştır.

Tablo 6: Angoff Yöntemi İçin Elde Edilen Kesme Puanlar

	Uzmanlar											Test
	U-1	U-2	U-3	U-4	U-5	U-6	U-7	U-8	U-9	U-10	U-11	
Kesme Puanı	15,10	12,80	14,15	13,40	9,90	14,50	10,60	10,40	17,15	14,65	12,55	13,20

Tablo 6’da uzman görüşleri doğrultusunda Angoff Yöntemi’ne ilişkin hesaplanan kesme puanlar yer almaktadır. İlk olarak söz konusu yönteme ilişkin uzmanların test maddelerine verdikleri puanlar (Ek - 4) toplanarak her bir uzmana ait kesme puan hesaplanmıştır. Bu doğrultuda uzmanlara ilişkin kesme puanlarının 9,9 ile 17,15 arasında değiştiği görülmektedir. Teste ilişkin kesme puanı ise uzmanlar için belirlenen kesme puanlarının aritmetik ortalaması alınarak hesaplanmıştır. Sonuç olarak Angoff Yöntemi doğrultusunda teste ilişkin nihai kesme puanı 13,2 bulunmuştur. Uzmanlar arası uyumu olumsuz etkilemeleri nedeniyle çalışma kapsamı dışında tutulan 3 uzmanın puanları dikkate alındığında testin kesme puanı 13,13 olarak hesaplanmıştır. Her iki kesme puanına göre başarılı sayılmak için gerekli olan minimum ham puanın 14,00 olması nedeniyle, çalışmadan çıkarılan uzmanların çalışma sonuçlarını etkilemediği görülmüştür.

Tablo 7: Nedelsky Yöntemi İçin Elde Edilen Kesme Puanlar

	Uzmanlar										Test
	U-1	U-2	U-3	U-4	U-5	U-6	U-7	U-8	U-9	U-10	
Kesme Puanı	7,48	9,65	8,42	6,08	15,58	9,50	7,98	6,78	5,28	8,42	8,52

Angoff Yöntemi’ne ilişkin kesme puanı hesaplandıktan sonra Nedelsky Yöntemi’ne geçilmiştir. Tablo 7’de uzman görüşleri doğrultusunda Nedelsky Yöntemi’ne ilişkin hesaplanan kesme puanlar yer almaktadır. İlk olarak uzmanların görüşleri doğrultusunda her bir test maddesine ilişkin elenen seçenek sayıları olasılık puanına (Ek - 5) dönüştürülmüştür. Bu süreçte Nedelsky Yöntemi’nde yer alan Tablo 1’den faydalanılmıştır. Elde edilen olasılıklar toplanarak Nedelsky Yöntemi’ne ilişkin her bir uzmana ait kesme puanlar hesaplanmıştır. Tablo 7’de uzmanlara ilişkin kesme puanların 5,28 ile 15,58 arasında değiştiği görülmektedir. Teste ilişkin kesme puanı ise uzmanlar için belirlenen kesme puanlarının aritmetik ortalaması alınarak hesaplanmıştır. Sonuç olarak Nedelsky Yöntemi doğrultusunda teste ilişkin nihai kesme puanı 8,52 bulunmuştur. Uzmanlar arası uyumu olumsuz etkilemeleri nedeniyle çalışma kapsamı dışında tutulan 4 uzmanın puanları dikkate alındığında testin kesme puanı 8,21 olarak

hesaplanmıştır. Her iki kesme puanına göre başarılı sayılmak için gerekli olan minimum ham puanın 9,00 olması nedeniyle, çalışmadan çıkarılan uzmanların çalışma sonuçlarını etkilemediği görülmüştür.

Standart belirleme yöntemlerine ilişkin kesme puanları belirlendikten sonra elde edilen kesme puanlar ve T puanlar doğrultusunda öğrencilerin başarı durumlarına karar verilmiştir. Tablo 8’de söz konusu kararlara ilişkin bilgiler yer almaktadır.

Tablo 8: Yöntemlere Ait Kesme Puanları, Başarılı Kabul Edilen Öğrenci Sayıları ve Yüzdeleri
(n = 107)

Yöntem	Değerlendirme	Kesme Puanı	Başarılı Kabul Edilen Öğrenci Sayısı	Başarılı Kabul Edilen Öğrenci Yüzdesi (%)
Angoff Y.	Ölçüt Dayanaklı	13,20	25	23
Nedelsky Y.	Ölçüt Dayanaklı	8,52	83	78
T Puanı	Norm Dayanaklı	50,00	50	47
T Puanı	Norm Dayanaklı	60,00	17	16

Tablo 8’e bakıldığında kullanılan yöntemlere göre alınan kararlarda farklılıklar görülmektedir. Kullanılan yöntem doğrultusunda başarılı olan öğrenci sayısı ve yüzdesi değişmektedir. Bu doğrultuda en az başarının olduğu yöntem, 60 T puanı ölçütünün kullanıldığı yöntem olan norm dayanaklı değerlendirmedir. Bu yöntemin dikkate alınması durumunda öğrencilerin %16’sının başarılı olduğu görülmektedir. Söz konusu yöntemi ölçüt dayanaklı değerlendirme olan Angoff Yöntemi takip etmektedir. Angoff Yöntemi’ne göre öğrencilerin %23’ünün başarılı olduğu görülmektedir. Angoff Yöntemi’ni %47 öğrenci başarı yüzdesi ile 50 T puanı ölçütünün kullanıldığı yöntem olan norm dayanaklı değerlendirme takip etmektedir. En yüksek başarının ise Nedelsky Yöntemi doğrultusunda elde edildiği görülmektedir. Bu yöntemin dikkate alınması durumunda öğrencilerin %78’inin başarılı olduğu görülmektedir. Nedelsky Yöntemi’ne göre başarı yüzdesinin diğer yöntemlerden yüksek çıkması, Çetin’in (2006) yaptığı çalışmayla örtüşmektedir.

Sonuç olarak kullanılan yöntemlere göre en düşük başarı yüzdesi ve en yüksek başarı yüzdesi arasındaki fark %62’dir. Kullanılan değerlendirme yöntemini değiştirmek, başarı yüzdesinde önemli bir değişime neden olmuştur. Bu doğrultuda ölçüt dayanaklı değerlendirme olan Nedelsky Yöntemi yerine norm dayanaklı değerlendirme yaparak 60 T puanı ölçütü dikkate alınır, testi alan öğrencilerin yarısından fazlasının başarı kararı değişecektir. Bu durum öğrencilerin başarıları değerlendirilirken yanlış kararlar vermeye neden olabilir. Söz konusu

farklılıkların anlamlılığı ise bağımlı iki oran arasındaki farkın testi ile alt problemlerin sırasına göre sınanmıştır.

3. 1. 1. Birinci Alt Probleme Ait Bulgular ve Yorumlar

Alt Problem: Angoff standart belirleme yöntemine ve 60 T puanı ölçütüne göre geçen öğrenci yüzdeleri arasında manidar fark var mıdır?

Birinci alt problemde öğrencilerin başarı yüzdeleri iki yöntemle göre belirlenmiştir. İlk yöntemde öğrencilerin T puanlarına göre 60 ölçütü dikkate alınarak başarı durumları belirlenmiştir. İkinci yöntemde ise Angoff Yöntemi doğrultusunda belirlenen 13,20 kesme puanı dikkate alınarak öğrencilerin başarı durumlarına ulaşılmıştır. İki yöntem doğrultusunda belirlenen başarı yüzdeleri arasındaki farka ilişkin sonuçlar ise Tablo 9’da sunulmuştur.

Tablo 9: Angoff Yöntemi’ne ve 60 T Puanı Ölçütüne İlişkin Başarılı Kabul Edilen Öğrenci Sayıları-Yüzdeleri ve z Değeri (n = 107)

Yöntem	Değerlendirme	Kesme Puanı	Başarılı Kabul Edilen Öğrenci Sayısı	Başarılı Kabul Edilen Öğrenci Yüzdesi (%)	z
Angoff Y.	Ölçüt Dayanaklı	13,20	25	23	2,83**
T Puanı	Norm Dayanaklı	60,00	17	16	

** p < 0,01

Tablo 9’a bakıldığında Angoff Yöntemi ve 60 T puanı ölçütü doğrultusunda öğrencilerin başarı yüzdeleri arasındaki fark dikkate alınarak hesaplanan z değeri anlamlı bulunmuştur. Bu durum, öğrenci başarısını belirlemede kullanılan söz konusu yöntemlerle öğrencilerin başarılı olma yüzdeleri arasında bir fark olduğunu göstermektedir. Bu doğrultuda 60 T puanı ölçütüne ve Angoff Yöntemi’ne göre geçen öğrenci yüzdeleri arasında 0,01 hata düzeyinde manidar bir fark olduğu sonucuna ulaşılmıştır.

Öğrencilerin başarı testi sonuçlarına göre Angoff Yöntemi dikkate alındığında başarı %23 ve 60 T puanı ölçütü dikkate alındığında başarı %16 olarak bulunmuştur. 60 T puanı ölçütüne göre başarı yüzdesiyle karşılaştırıldığında Angoff Yöntemi’nde başarının %7 daha fazla olduğu görülmektedir. Sadece kullanılan değerlendirme yöntemini değiştirmek öğrencilerin başarı yüzdelerinde farklılaşmaya neden olmuştur. Norm dayanaklı değerlendirme yaparken ham puanların T puanına dönüştürülmesi ve 60 T puanı ölçütünün Angoff Yöntemi’ndeki kesme

puanından farklı bir ham puana denk gelmesi, yüzdelerdeki farklılaşmaya neden olarak gösterilebilir.

3. 1. 2. İkinci Alt Probleme Ait Bulgular ve Yorumlar

Alt Problem: Angoff standart belirleme yöntemine ve 50 T puanı ölçütüne göre geçen öğrenci yüzdeleri arasında manidar fark var mıdır?

İkinci alt problemde öğrencilerin başarı yüzdeleri iki yöneme göre belirlenmiştir. İlk yöntemde öğrencilerin T puanlarına göre 50 ölçütü dikkate alınarak başarı durumları belirlenmiştir. İkinci yöntemde ise Angoff Yöntemi doğrultusunda belirlenen 13,20 kesme puanı dikkate alınarak öğrencilerin başarı durumlarına ulaşılmıştır. İki yöntem doğrultusunda belirlenen başarı yüzdeleri arasındaki farka ilişkin sonuçlar ise Tablo 10'da sunulmuştur.

Tablo 10: Angoff Yöntemi'ne ve 50 T Puanı Ölçütüne İlişkin Başarılı Kabul Edilen Öğrenci Sayıları-Yüzdeleri ve z Değeri (n = 107)

Yöntem	Değerlendirme	Kesme Puanı	Başarılı Kabul Edilen Öğrenci Sayısı	Başarılı Kabul Edilen Öğrenci Yüzdesi (%)	z
Angoff Y.	Ölçüt Dayanaklı	13,20	25	23	-5,00**
T Puanı	Norm Dayanaklı	50,00	50	47	

** p < 0,01

Tablo 10'a bakıldığında Angoff Yöntemi ve 50 T puanı ölçütü doğrultusunda öğrencilerin başarı yüzdeleri arasındaki fark dikkate alınarak hesaplanan z değeri anlamlı bulunmuştur. Bu doğrultuda 50 T puanı ölçütüne ve Angoff Yöntemi'ne göre geçen öğrenci yüzdeleri arasında 0,01 hata düzeyinde manidar bir fark olduğu sonucuna ulaşılmıştır. Bu durum, öğrenci başarısını belirlemede kullanılan değerlendirme yöntemlerinin öğrencilerin başarı yüzdelerini farklılaştırdığını göstermektedir.

Öğrencilerin başarı testi sonuçlarına göre Angoff Yöntemi dikkate alındığında başarı %23 ve 50 T puanı ölçütü dikkate alındığında başarı %47 olarak bulunmuştur. Ölçüt dayanaklı değerlendirme olan Angoff Yöntemi'ne göre başarı yüzdesiyle karşılaştırıldığında norm dayanaklı değerlendirme olan 50 ölçütüne göre başarının %24 daha fazla olduğu görülmektedir. Bu sonuç Nartgün'ün (2007) çalışma sonuçlarıyla örtüşmektedir. Kullanılan değerlendirme yöntemi doğrultusunda başarılı olan öğrencilerin yüzdeleri değişmiştir. Angoff Yöntemi'ndeki

kesme puanının 50 T puanı ölçütüne denk gelen ham puandan daha yüksek olması bu değişime neden olarak gösterilebilir.

3. 1. 3. Üçüncü Alt Probleme Ait Bulgular ve Yorumlar

Alt Problem: Nedelsky standart belirleme yöntemine ve 60 T puanı ölçütüne göre geçen öğrenci yüzdeleri arasında manidar fark var mıdır?

Üçüncü alt problemde öğrencilerin başarı yüzdeleri iki yönetime göre belirlenmiştir. İlk yöntemde öğrencilerin T puanlarına göre 60 ölçütü dikkate alınarak başarı durumları belirlenmiştir. İkinci yöntemde ise Nedelsky Yöntemi doğrultusunda belirlenen 8,52 kesme puanı dikkate alınarak öğrencilerin başarı durumlarına ulaşılmıştır. İki yöntem doğrultusunda belirlenen başarı yüzdeleri arasındaki farka ilişkin sonuçlar ise Tablo 11’de sunulmuştur.

Tablo 11: Nedelsky Yöntemi’ne ve 60 T Puanı Ölçütüne İlişkin Başarılı Kabul Edilen Öğrenci Sayıları-Yüzdeleri ve z Değeri (n = 107)

Yöntem	Değerlendirme	Kesme Puanı	Başarılı Kabul Edilen Öğrenci Sayısı	Başarılı Kabul Edilen Öğrenci Yüzdesi (%)	z
Nedelsky Y.	Ölçüt Dayanaklı	8,52	83	78	8,12**
T Puanı	Norm Dayanaklı	60,00	17	16	

** p < 0,01

Tablo 11’e bakıldığında Nedelsky Yöntemi ve 60 T puanı ölçütü doğrultusunda öğrencilerin başarı yüzdeleri arasındaki fark dikkate alınarak hesaplanan z değeri anlamlı bulunmuştur. Bu doğrultuda 60 T puanı ölçütüne ve Nedelsky Yöntemi’ne göre geçen öğrenci yüzdeleri arasında 0,01 hata düzeyinde manidar bir fark olduğu sonucuna ulaşılmıştır. Bu durum, öğrenci başarısını belirlemede kullanılan söz konusu yöntemlerle öğrencilerin başarı yüzdeleri arasında bir ilişki olduğunu göstermektedir.

Öğrencilerin başarı testi sonuçlarına göre Nedelsky Yöntemi dikkate alındığında başarı %78 ve 60 T puanı ölçütü dikkate alındığında başarı %16 olarak bulunmuştur. Nedelsky Yöntemi’ne göre başarı yüzdesiyle karşılaştırıldığında 60 T puanı ölçütüne göre başarının %62 daha az olduğu görülmektedir. Kullanılan iki farklı değerlendirme yöntemi doğrultusunda öğrencilerin başarı yüzdelerinde büyük bir fark olduğu bulunmuştur. Değerlendirme yöntemlerinden her birinin kullandığı ölçütün ve kesme puanının birbirinden çok farklı olması öğrencilerin başarı yüzdelerindeki farklılaşmaya neden olduğu söylenebilir. Bu sonuçlar Oescher ve diğerlerinin

(1992) bulduğu sonuçlardan farklılık göstermektedir. Oescher'in çalışmasında farklı alanlara ilişkin başarının değerlendirilmesi ve çalışma grubunun bu araştırmadakinden daha geniş olması bu farklılığa neden olarak görülebilir.

3. 1. 4. Dördüncü Alt Probleme Ait Bulgular ve Yorumlar

Alt Problem: Nedelsky standart belirleme yöntemine ve 50 T puanı ölçütüne göre geçen öğrenci yüzdeleri arasında manidar fark var mıdır?

Dördüncü alt problemde öğrencilerin başarı yüzdeleri iki yöntemle göre belirlenmiştir. İlk yöntemde öğrencilerin T puanlarına göre 50 ölçütü dikkate alınarak başarı durumları belirlenmiştir. İkinci yöntemde ise Nedelsky Yöntemi doğrultusunda belirlenen 8,52 kesme puanı dikkate alınarak öğrencilerin başarı durumlarına ulaşılmıştır. İki yöntem doğrultusunda belirlenen başarı yüzdeleri arasındaki farka ilişkin sonuçlar ise Tablo 12'de sunulmuştur.

Tablo 12: Nedelsky Yöntemi'ne ve 50 T Puanı Ölçütüne İlişkin Başarılı Kabul Edilen Öğrenci Sayıları-Yüzdeleri ve z Değeri (n = 107)

Yöntem	Değerlendirme	Kesme Puanı	Başarılı Kabul Edilen Öğrenci Sayısı	Başarılı Kabul Edilen Öğrenci Yüzdesi (%)	z
Nedelsky Y.	Ölçüt Dayanaklı	8,52	83	78	5,74**
T Puanı	Norm Dayanaklı	50,00	50	47	

** p < 0,01

Tablo 12'ye bakıldığında Nedelsky Yöntemi ve 50 T puanı ölçütü doğrultusunda öğrencilerin başarı yüzdeleri arasındaki fark dikkate alınarak hesaplanan z değeri anlamlı bulunmuştur. Bu doğrultuda 50 T puanı ölçütüne ve Nedelsky Yöntemi'ne göre geçen öğrenci yüzdeleri arasında 0,01 hata düzeyinde manidar bir fark olduğu sonucuna ulaşılmıştır. Bu durum, öğrenci başarısını belirlemede kullanılan farklı değerlendirme yöntemlerine göre başarıya ilişkin farklı kararlara ulaşılabildiğini göstermektedir.

Öğrencilerin başarı testi sonuçlarına göre Nedelsky Yöntemi dikkate alındığında başarı %78 ve 50 T puanı ölçütü dikkate alındığında başarı %47 olarak bulunmuştur. 50 T puanı ölçütüne göre başarı yüzdesiyle karşılaştırıldığında Nedelsky Yöntemi'nde başarının %31 daha fazla olduğu görülmektedir. Öğrencilerin başarı testi sonuçlarını değerlendirirken kullanılan karar ölçütü, başarılı kabul edilen öğrencilerin yüzdesinde önemli bir değişime neden olmuştur.

Değerlendirme yöntemlerine ilişkin minimum başarı düzeylerinin farklı olması, öğrenci başarısına ilişkin yüzdelerde farklılaşmaya neden olarak gösterilebilir.

3. 1. 5. Beşinci Alt Probleme Ait Bulgular ve Yorumlar

Alt Problem: 50 ve 60 T puanlarına göre geçen öğrenci yüzdeleri arasında manidar fark var mıdır?

Beşinci alt problemde öğrencilerin başarı yüzdeleri iki yönetime göre belirlenmiştir. İlk yöntemde öğrencilerin T puanlarına göre 50 ölçütü dikkate alınarak başarı durumları belirlenmiştir. İkinci yöntemde benzer şekilde öğrencilerin T puanları 60 ölçütü dikkate alınarak başarı durumlarına ulaşılmıştır. İki yöntem doğrultusunda belirlenen başarı yüzdeleri arasındaki farka ilişkin sonuçlar ise Tablo 13’de sunulmuştur.

Tablo 13: 50 ve 60 T Puanı Ölçütlerine İlişkin Başarılı Kabul Edilen Öğrenci Sayıları- Yüzdeleri ve z Değeri (n = 107)

Yöntem	Değerlendirme	Kesme Puanı	Başarılı Kabul Edilen Öğrenci Sayısı	Başarılı Kabul Edilen Öğrenci Yüzdesi (%)	z
T Puanı	Norm Dayanaklı	50,00	50	47	-5,74**
T Puanı	Norm Dayanaklı	60,00	17	16	

** p < 0,01

Tablo 13’e bakıldığında 50 ve 60 T puanı ölçütleri doğrultusunda öğrencilerin başarı yüzdeleri arasındaki fark dikkate alınarak hesaplanan z değeri anlamlı bulunmuştur. Bu durum, öğrenci başarısını belirlemede kullanılan söz konusu yöntemlerle öğrencilerin başarı yüzdeleri arasında bir ilişki olduğunu göstermektedir. Bu doğrultuda 50 ve 60 T puanlarına göre geçen öğrenci yüzdeleri arasında 0,01 hata düzeyinde manidar bir fark olduğu görülmektedir. Öğrencilerin başarı testi sonuçlarına göre 50 T puan ölçüt olarak dikkate alındığında başarı %47 ve 60 T puanı ölçüt olarak dikkate alındığında başarı %16 olarak bulunmuştur. Kullanılan bağıl ölçütü 10 puan yukarıya çekmek öğrenci başarısını %31 düşürmüştür.

Öğrencilerin başarı testi sonuçlarını bağıl ölçüt kullanılarak değerlendirirken kullanılan karar ölçütünü değiştirmek, başarılı kabul edilen öğrencilerin yüzdesinde önemli bir değişime neden olmuştur. Ancak başarının normal dağılıma sahip olduğu düşünüldüğünde ölçütte 10 puanlık artış 1 standart sapmaya karşılık gelmektedir. T puanında ise ortalamanın 50 olması nedeniyle 1 standart sapma başarının artırılması başarı kararında %34’lük bir düşüşe denk gelmektedir. Bu

nedenle T puanının 50'den 60'a artırılması nedeniyle başarıda %34'lük bir düşüş beklenen bir durumdur. Çalışmada, başarı yüzdesindeki düşüşe bu durum neden olarak gösterilebilir.

Sonuç olarak literatürde yer alan birçok araştırmanın aksine bu çalışma kapsamında kullanılan yöntemlerin tamamı doğrultusunda norm ve ölçüt dayanaklı değerlendirmelerin geçme yüzdeleri arasında manidar farklılıklara ulaşılmıştır. Ayrıca 50 ve 60 T puanı ölçütlerine göre de başarı yüzdelerinin değiştiği sonucuna ulaşılmıştır. Bu doğrultuda kullanılan değerlendirme yöntemine göre başarılı olan öğrenci yüzdeleri arasında önemli farklılıklar vardır. Nartgün'ün (2007) çalışmasında, norm ve ölçüt dayanaklı değerlendirmeler arası bulunan farklar ile bu çalışma sonuçları örtüşmektedir.

3. 2. İKİNCİ PROBLEME AİT BULGULAR VE YORUMLAR

Problem: Norm ve ölçüt dayanaklı değerlendirmelerin uygunluk geçerlikleri nasıldır?

Norm ve ölçüt dayanaklı değerlendirmeler doğrultusunda alınan kararların uygunluk geçerlikleri incelenirken çalışmaya katılan öğrencilerin kendi okullarında ölçme ve değerlendirme dersinden dönem sonunda aldıkları notlar ölçüt olarak kullanılmıştır. Dönem sonu notları ve bu çalışmadan elde edilen puanlar başarılı ve başarısız olmak üzere iki kategoride sınıflanmıştır. Bu nedenle araştırma kapsamında yer alan yöntemlere ilişkin uygunluk geçerlikleri alt problemlerin sırasına göre Cohen'in Kappa testi ile incelenmiştir.

3. 2. 1. Birinci Alt Probleme Ait Bulgular ve Yorumlar

Alt Problem: 60 T puanı ölçütüne göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyum nasıldır?

Birinci alt problemde, çalışmada uygulanan teste ilişkin başarı durumları 60 T puanı ölçütüne göre belirlenerek öğrencilerin kendi okullarındaki ölçme ve değerlendirme dersine ilişkin başarı durumlarıyla uyumları incelenmiştir. İki farklı gözlem sonuçları arasındaki uyumu gösteren Cohen'in Kappa Testi sonuçları Tablo 14'de sunulmuştur.

Tablo 14: 60 T Puanı Ölçütüne İlişkin Başarı Durumu ve Dönem Sonu Ders Başarısı Durumu Arasındaki Uyum Frekansı, Yüzdesi ve Kappa Değeri (n=107)

Uyum Frekansı	Uyum Yüzdesi	Kappa
38	35,51	0,09*

* p < 0,05

Tablo 14'e bakıldığında öğrencilerin 60 T puanı ölçütüne ilişkin başarı durumu ve dönem sonu ders başarısı durumu arasındaki uyum dikkate alınarak hesaplanan Kappa değerinin 0,05 hata düzeyinde manidar olduğu görülmektedir. Ancak Kappa değeri, düşük düzeyde ($Kappa < 0,40$) bulunmuştur. Bu nedenle 60 T puanı ölçütüne ilişkin alınan kararların dış ölçüt olan dönem sonu ders notlarıyla uyumunun düşük olduğu bulunmuştur. Yüzdelere arasındaki uyum incelendiğinde ise iki farklı gözleme ilişkin başarı sınıflamasında %35,51 uyum olduğu sonucuna ulaşılmıştır. 69 öğrenciye denk gelen %64,49'luk başarıya ilişkin kararın ise iki farklı gözleme göre uyumsuz olduğu görülmüştür. Bu doğrultuda 60 T puanı ölçütüne ilişkin kararların uygunluk geçerliğinin düşük olduğu sonucuna ulaşılmıştır. Uygunluk geçerliği dikkate alındığında 60 T puanı ölçütüne ilişkin alınan kararların yeterince başarılı ve geçerli olmadığı görülmüştür. Öğrencilerin ders notlarının yüksek olması ve çalışmada uygulanan teste 60 T puanı ölçütüne göre öğrenci başarılarının düşük olması, kararlar arasındaki düşük uyuma neden gösterilebilir. Ayrıca 60 T puanı ölçütünün norm dayanaklı değerlendirme olması ve bu yöntemin testin uygulanma amacı için uygun olmaması, kararlardaki uyumsuzluğu etkilemiş olabilir.

3. 2. 2. İkinci Alt Probleme Ait Bulgular ve Yorumlar

Alt Problem: 50 T puanı ölçütüne göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyum nasıldır?

İkinci alt problemde, araştırmada uygulanan teste ilişkin başarı durumları 50 T puanı ölçütüne göre belirlenerek öğrencilerin kendi okullarındaki ölçme ve değerlendirme dersine ilişkin başarı durumlarıyla uyumları incelenmiştir. İki farklı gözlem sonuçları arasındaki uyumu gösteren Cohen'in Kappa Testi sonuçları Tablo 15'de sunulmuştur.

Tablo 15: 50 T Puanı Ölçütüne İlişkin Başarı Durumu ve Dönem Sonu Ders Başarısı Durumu Arasındaki Uyum Frekansı, Yüzdesi ve Kappa Değeri (n=107)

Uyum Frekansı	Uyum Yüzdesi	Kappa
65	60,75	0,24**

** $p < 0,01$

Tablo 15'e bakıldığında öğrencilerin 50 T puanı ölçütüne ilişkin başarı durumu ve dönem sonu ders başarısı durumu arasındaki uyum dikkate alınarak hesaplanan Kappa değerinin 0,01 hata düzeyinde manidar olduğu görülmektedir. Ancak Kappa değeri, düşük düzeyde ($Kappa < 0,40$) bulunmuştur. Bu nedenle 50 T puanı ölçütüne ilişkin alınan kararların dış ölçüt olan dönem

sonu ders notlarıyla uyumunun düşük olduğu bulunmuştur. Yüzdeler arasındaki uyum incelendiğinde ise iki farklı gözleme ilişkin başarı sınıflamasında %60,75 uyum olduğu sonucuna ulaşılmıştır. 42 öğrenciye denk gelen %39,25'lik başarıya ilişkin kararın ise iki farklı gözleme göre uyumsuz olduğu görülmüştür. Yüzdesel uyumun Kappa değerine yeterince yansımaması, Kappa değerinin şans faktörünü dikkate almasıyla açıklanabilir. Bu doğrultuda 50 T puanı ölçütüne ilişkin kararların uygunluk geçerliğinin düşük olduğu sonucuna ulaşılmıştır. Uygunluk geçerliği dikkate alındığında 50 T puanı ölçütüne ilişkin alınan kararların yeterince başarılı ve geçerli olmadığı görülmüştür. 50 T puanı ölçütünün norm dayanaklı değerlendirme olması ve bu yöntemin testin uygulanma amacı için uygun olmaması, kararlardaki uyumsuzluğu etkilemiş olabilir.

3. 2. 3. Üçüncü Alt Probleme Ait Bulgular ve Yorumlar

Alt Problem: Angoff standart belirleme yöntemine göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyum nasıldır?

Üçüncü alt problemde, araştırmada uygulanan teste ilişkin başarı durumları Angoff Yöntemi'ne göre belirlenerek öğrencilerin kendi okullarındaki ölçme ve değerlendirme dersine ilişkin başarı durumlarıyla uyumları incelenmiştir. İki farklı gözlem sonuçları arasındaki uyumu gösteren Cohen'in Kappa Testi sonuçları Tablo 16'da sunulmuştur.

Tablo 16: Angoff Yöntemi'ne İlişkin Başarı Durumu ve Dönem Sonu Ders Başarısı Durumu Arasındaki Uyum Frekansı, Yüzdesi ve Kappa Değeri (n=107)

Uyum Frekansı	Uyum Yüzdesi	Kappa
46	42,99	0,14**

** p < 0,01

Tablo 16'ya bakıldığında öğrencilerin Angoff Yöntemi'ne ilişkin başarı durumu ve dönem sonu ders başarısı durumu arasındaki uyum dikkate alınarak hesaplanan Kappa değerinin 0,01 hata düzeyinde manidar olduğu görülmektedir. Ancak Kappa değeri, düşük düzeyde (Kappa<0,40) bulunmuştur. Bu nedenle Angoff Yöntemi'ne ilişkin alınan kararların dış ölçüt olan dönem sonu ders notlarıyla uyumunun düşük olduğu bulunmuştur. Yüzdeler arasındaki uyum incelendiğinde ise iki farklı gözleme ilişkin başarı sınıflamasında %42,99 uyum olduğu sonucuna ulaşılmıştır. 61 öğrenciye denk gelen %57,01'lik başarıya ilişkin kararın ise iki farklı gözleme göre uyumsuz olduğu görülmüştür. Bu doğrultuda ölçüt dayanaklı değerlendirme olan Angoff Yöntemi'ne ilişkin kararların uygunluk geçerliğinin düşük olduğu sonucuna ulaşılmıştır. Uygunluk geçerliği

dikkate alındığında Angoff Yöntemi'ne ilişkin alınan kararların yeterince başarılı ve geçerli olmadığı görülmüştür. Öğrencilerin ders notlarının yüksek olması ve çalışmada uygulanan testte Angoff Yöntemi'ne göre öğrenci başarılarının düşük olması, kararlar arasındaki düşük uyuma neden gösterilebilir.

3. 2. 4. Dördüncü Alt Probleme Ait Bulgular ve Yorumlar

Alt Problem: Nedelsky standart belirleme yöntemine göre geçme ve kalma kararları ile dönem sonunda söz konusu dersten geçme ve kalma kararları arasındaki uyum nasıldır?

Dördüncü alt problemde, araştırmada uygulanan teste ilişkin başarı durumları Nedelsky Yöntemi'ne göre belirlenerek öğrencilerin kendi okullarındaki ölçme ve değerlendirme dersine ilişkin başarı durumlarıyla uyumları incelenmiştir. İki farklı gözlem sonuçları arasındaki uyumu gösteren Cohen'in Kappa Testi sonuçları Tablo 17'de sunulmuştur.

Tablo 17: Nedelsky Yöntemi'ne İlişkin Başarı Durumu ve Dönem Sonu Ders Başarısı Durumu Arasındaki Uyum Frekansı, Yüzdesi ve Kappa Değeri (n=107)

Uyum Frekansı	Uyum Yüzdesi	Kappa
86	80,37	0,41**

** p < 0,01

Tablo 17'ye bakıldığında öğrencilerin Nedelsky Yöntemi'ne ilişkin başarı durumu ve dönem sonu ders başarısı durumu arasındaki uyum dikkate alınarak hesaplanan Kappa değerinin 0,01 hata düzeyinde manidar olduğu görülmektedir. Bu doğrultuda Kappa değerinin iyi düzeyde uyumlu olduğu ($Kappa > 0,40$) bulunmuştur. Bu nedenle Nedelsky Yöntemi'ne ilişkin alınan kararların dış ölçüt olan dönem sonu ders notlarıyla uyumlu olduğu bulunmuştur.

Yüzdeler arasındaki uyum incelendiğinde ise iki farklı gözleme ilişkin başarı sınıflamasında %80,37 uyum olduğu sonucuna ulaşılmıştır. 21 öğrenciye denk gelen %19,63'lük başarıya ilişkin kararın ise iki farklı gözleme göre uyumsuz olduğu bulunmuştur. Sonuç olarak ölçüt dayanaklı değerlendirme olan Nedelsky Yöntemi'ne ilişkin kararların uygunluk geçerliğine sahip olduğu bulunmuştur. Uygunluk geçerliği dikkate alındığında Nedelsky Yöntemi'ne ilişkin alınan kararların başarılı ve geçerli olduğu görülmüştür.

Öğrencilerin ders notları dikkate alındığında başarı oranı yüksektir. Araştırmada uygulanan test sonuçları Nedelsky Yöntemi'ne göre değerlendirildiğinde ise ders notlarındaki başarıya benzer şekilde öğrencilerin başarı yüzdesi yüksektir. Bu durum ders notlarıyla Nedelsky Yöntemi'ne

ilişkin kararların uyumlu olmasına neden olarak gösterilebilir. Ayrıca Nedelsky Yöntemi'nin ölçüt dayanaklı değerlendirme olması ve bu yöntemin testin uygulanma amacı için daha uygun olması, kararlardaki uyumu etkilemiş olabilir. Araştırma kapsamındaki diğer ölçüt dayanaklı değerlendirme olan Angoff Yöntemi'ne oranla Nedelsky Yöntemi'nin uygunluk geçerliğinin yüksek olması ise yöntemlerin uygulama süreçlerindeki farklılıktan kaynaklanıyor olabilir.

Ölçüt olarak kullanılan öğrencilerin dersten başarılı veya başarısız olma durumları ile çalışma kapsamındaki yöntemlere ilişkin başarılı veya başarısız olma durumları arasındaki uyum, yöntemden yönteme farklılık göstermiştir. Uyum yüzdeleri ve Kappa değerleri doğrultusunda sadece ölçüt dayanaklı değerlendirme olan Nedelsky Yöntemi'nin uygunluk geçerliğine sahip olduğu sonucuna ulaşılmıştır. Öğrencilerin ders başarılarının yüksek olması ve Nedelsky Yöntemi'nin başarı ölçütünü düşük tutması buna neden olarak gösterilebilir. Norm dayanaklı değerlendirme yöntemlerinin (50 ve 60 T puanları) ve ölçüt dayanaklı değerlendirmelerden Angoff Yöntemi'nin Kappa değeri doğrultusunda uygunluk geçerliklerinin düşük olduğu bulunmuştur. Buna neden olarak ise söz konusu yöntemlerin başarı ölçütünün Nedelsky Yöntemi'ne oranla daha yüksek olması gösterilebilir. Yüzdesele uyum açısından ise 50 T puanı ölçütüne ilişkin kararların uygunluk geçerliğinin 60 T puanı ölçütüne oranla daha yüksek olduğu gözlemlenmiştir.

Elde edilen sonuçlar dikkate alındığında Jacobson (2008) tarafından yapılan çalışma sonuçlarından farklı sonuçlara ulaşılmıştır. Jacobson, norm dayanaklı değerlendirmelerle ölçüt dayanaklı değerlendirmelerin öğrencileri yeterli ve yetersiz olarak iki kategoride başarılı bir şekilde sınıfladığı sonucuna ulaşmıştır. Bu çalışmada ise norm ve ölçüt dayanaklı değerlendirmelerle alınan kararların geçerliklerinin yöntemden yönteme değiştiği bulunmuştur. Uygunluk geçerliği açısından en başarılı yöntemin ise ölçüt dayanaklı değerlendirme olan Nedelsky Yöntemi olduğu sonucuna ulaşılmıştır. Jacobson'un ve bu araştırmanın sonuçlarındaki farklılık, çalışmalarda kullanılan testlerin kapsam alanlarındaki farklılıktan kaynaklanıyor olabilir. Ayrıca bu çalışmada bir test üzerinden değerlendirme yapılırken Jacobson'un norm ve ölçüt dayanaklı değerlendirmeler için farklı iki test uygulaması da bu farklılığa neden olarak gösterilebilir. Bunun yanında Jacobson'un başarı seviyesini belirlerken görüşüne başvurduğu uzman sayısının bu çalışmadakinden fazla olması da sonuçların farklılaşmasına neden olarak gösterilebilir.

BÖLÜM IV

SONUÇ VE ÖNERİLER

Bu bölümde, araştırmadan elde edilen bulgular doğrultusunda sonuçlar belirtilmiş ve sonuçlara dayalı olarak kurumlara ve araştırmacılara bazı önerilerde bulunulmuştur.

4. 1. SONUÇLAR

Araştırmadan elde edilen sonuçlar, problemlerin sırasına göre değinilmiştir. Alt problemlere ilişkin sonuçlar ise ait oldukları problem çerçevesinde sıralanmıştır.

4. 1. 1. Birinci Probleme Ait Sonuçlar

1. Ölçüt dayanaklı değerlendirme olan Angoff Yöntemi'ne göre alınan kararlar ile norm dayanaklı değerlendirmede 60 ölçütüne göre alınan kararlar farklılık göstermiştir. Araştırmada Angoff Yöntemi'ne göre başarı yüzdesi, 60 T puanı ölçütüne göre başarı yüzdesinden daha yüksek çıkmıştır. Angoff Yöntemi ve 60 T puanı ölçütüne göre geçen öğrenci yüzdeleri manidar düzeyde farklılık göstermiştir. Araştırmada kullanılan yöntemlere ilişkin en düşük düzeydeki farklılaşma, Angoff Yöntemi ve 60 T puanı ölçütü arasında görülmüştür.

2. Ölçüt dayanaklı değerlendirme olan Angoff Yöntemi'ne göre alınan kararlar ile norm dayanaklı değerlendirmede 50 ölçütüne göre alınan kararlar farklılık göstermiştir. Araştırmada 50 T puanı ölçütüne göre başarı yüzdesi, Angoff Yöntemi'ne göre başarı yüzdesinden daha yüksek çıkmıştır. Angoff Yöntemi ve 50 T puanı ölçütüne göre geçen öğrenci yüzdeleri manidar düzeyde farklılık göstermiştir.

3. Ölçüt dayanaklı değerlendirme olan Nedelsky Yöntemi'ne göre alınan kararlar ile norm dayanaklı değerlendirmede 60 ölçütüne göre alınan kararlar farklılaşmıştır. Araştırmada Nedelsky Yöntemi'ne göre başarı yüzdesi, 60 T puanı ölçütüne göre başarı yüzdesinden daha yüksek çıkmıştır. Nedelsky Yöntemi ve 60 T puanı ölçütüne göre geçen öğrenci yüzdeleri manidar düzeyde farklılık göstermiştir. Araştırmada kullanılan yöntemlere ilişkin en yüksek düzeydeki farklılaşma, Nedelsky Yöntemi ve 60 T puanı ölçütü arasında görülmüştür.

4. Ölçüt dayanaklı değerlendirme olan Nedelsky Yöntemi'ne göre alınan kararlar ile norm dayanaklı değerlendirmede 50 ölçütüne göre alınan kararlar farklılaşmıştır. Araştırmada Nedelsky Yöntemi'ne göre başarı yüzdesi, 50 T puanı ölçütüne göre başarı yüzdesinden daha

yüksek çıkmıştır. Nedelsky Yöntemi ve 50 T puanı ölçütüne göre geçen öğrenci yüzdeleri manidar düzeyde farklılık göstermiştir.

5. 50 ve 60 T puanlarına göre öğrenci başarılarına ilişkin alınan kararların farklılık gösterdiği sonucuna varılmıştır. 50 ve 60 T puanlarına göre geçen öğrenci yüzdeleri manidar düzeyde farklılık göstermiştir.

4. 1. 2. İkinci Probleme Ait Sonuçlar

1. 60 T puanı ölçütüne ilişkin başarı kararları ile dönem sonu ders başarısına ilişkin kararlar arasındaki uyumun manidar olduğu görülmüştür. İki yöntem arasındaki uyumun düşük düzeyde olduğu görülmüştür. Bu doğrultuda 60 T puanı ölçütüne ilişkin kararların uygunluk geçerliğinin düşük olduğu sonucuna varılmıştır.

2. 50 T puanı ölçütüne ilişkin başarı kararları ile dönem sonu ders başarısına ilişkin kararlar arasındaki uyumun manidar olduğu görülmüştür. İki yöntem arasındaki uyumun düşük düzeyde olduğu görülmüştür. Bu doğrultuda 50 T puanı ölçütüne ilişkin kararların uygunluk geçerliğinin düşük olduğu sonucuna varılmıştır.

3. Ölçüt dayanaklı değerlendirme olan Angoff Yöntemi'ne ilişkin başarı kararları ile dönem sonu ders başarısına ilişkin kararlar arasındaki uyumun manidar olduğu görülmüştür. İki yöntem arasındaki uyumun düşük düzeyde olduğu görülmüştür. Bu doğrultuda Angoff Yöntemi'ne ilişkin kararların uygunluk geçerliğinin düşük olduğu sonucuna varılmıştır.

4. Ölçüt dayanaklı değerlendirme olan Nedelsky Yöntemi'ne ilişkin başarı kararları ile dönem sonu ders başarısına ilişkin kararlar arasındaki uyumun manidar olduğu görülmüştür. İki yöntem arasındaki uyumun iyi olduğu görülmüştür. Bu doğrultuda Nedelsky Yöntemi'ne ilişkin kararların uygunluk geçerliğine sahip olduğu sonucuna varılmıştır.

4. 2. ÖNERİLER

4. 2. 1. Araştırma Sonuçlarından Çıkan Öneriler

- Kullanılan değerlendirme yöntemine göre öğrencilerin başarı yüzdelerinde farklılaşma olacağı için kurumların amaçlarına uygun değerlendirme yöntemini kullanmaları önerilir. Bu doğrultuda özellikle kişilerin performans seviyelerinin veya yeterlilik durumlarının belirleneceği değerlendirmelerde ölçüt dayanaklı değerlendirme kullanılması önerilir.

- Nedelsky Yöntemi ile yapılan değerlendirmenin ölçüt dayanaklı olması ve bu çalışma sonucuna göre uygunluk geçerliğine sahip olması nedeniyle özellikle eğitim kurumlarının, bireylerin bir alana ilişkin yeterliliklerini değerlendirirken yeterli sayıda uzmana ulaşılabilecek ve yeterli zamana sahip olunabilecek durumlarda Nedelsky Yöntemi'ni kullanmaları önerilir.
- Yüzdesel uyum doğrultusunda 50 T Puanı ölçütünün uygunluk geçerliğinin 60 T puanından daha yüksek olması nedeniyle, norm dayanaklı değerlendirme yapılması gereken durumlarda 60 T puanı ölçütü yerine 50 T puanı ölçütü kullanılabilir.

4. 2. 2. Araştırmacılara Yönelik Öneriler

- Angoff ve Nedelsky Yöntemleri dışındaki diğer standart belirleme yöntemleri dikkate alınarak benzer bir çalışma yapılabilir. Bu doğrultuda özellikle birey odaklı standart belirleme yöntemleri ile norm dayanaklı değerlendirmeler karşılaştırılabilir.
- Yüksek lisans ve doktora programlarında kullanılan minimum başarı puanları değişebilmektedir. Bu doğrultuda 50 ve 60 T puanı ölçütleri dışında kullanılan minimum başarı ölçütleri ile standart belirleme yöntemleriyle hesaplanan kesme puanları dikkate alınarak öğrencilerin başarı yüzdeleri karşılaştırılabilir.
- Ölçme ve değerlendirme dışındaki başka dersler için de benzer bir çalışma yapılabilir. Bu doğrultuda norm ve ölçüt dayanaklı değerlendirmelere ilişkin öğrencilerin başarı yüzdeleri, dersin sayısal ya da sözel içerikli olmasına göre değişip değişmediği incelenebilir.
- Norm olarak çalışma grubunun verilerini kullanmak yerine çalışmanın yapılacağı derse ilişkin uygun bir norm grup oluşturularak bu grubun ortalama ve standart sapma değerleri doğrultusunda çalışma grubunun puanları T puanına dönüştürülebilir. Bu sayede norm ve ölçüt dayanaklı değerlendirmeler arasındaki farklar araştırılabilir.
- Benzer bir çalışma yordama geçerliği kullanılarak yapılabilir. Bu doğrultuda norm ve ölçüt dayanaklı değerlendirmelere ilişkin alınan kararların gelecekteki bir durumu ne ölçüde yordadıkları karşılaştırılabilir. Örneğin ortaokul öğrencilerini kapsayan bir çalışmada herhangi bir derse ilişkin farklı yöntemlerle alınan kararların liselere giriş sınavında alınan puanları ne ölçüde yordadığı araştırılabilir.
- Benzer bir çalışma eğitim dışındaki alanlarda da yapılabilir. Özellikle satış ve pazarlama gibi alanlarda performans gibi daha amaca uygun dış ölçütler bulunabilmektedir. Bu doğrultuda satış ve pazarlama alanına ilişkin yetenek testi

sonuları farklı yöntemlerle deęerlendirilebilir ve alınan kararların geerlięini incelemek iin alıřmaya katılan bireylerin satıř performanslarıyla karřılařtırılabilir.

- Bu alıřmada ęrenciler bařarılı ve bařarısız olmak üzere iki kategoride sınıflandırılmıř ve bu doęrultuda norm dayanaklı deęerlendirmelerle lüt dayanaklı deęerlendirmeler karřılařtırılmıřtır. Bařarı sınıflamasındaki kategori sayısı artırılarak benzer bir alıřma yapılabilir. rneęin üniversitelerde kullanılan harf sistemine gre kategori sayısı belirlenebilir ya da Jacobson (2008) tarafından yapılan alıřmadaki gibi bařarı drt kategoride incelenebilir.

KAYNAKÇA

- American Federation of Teachers, National Council on Measurement in Education ve National Education Association. (1990). *Standards for teacher competence in educational assessment of students*. Erişim: 07 Ocak 2013, <http://www.ode.state.or.us>
- Aiken, L. R. (2000). *Psychological testing and assessment*. Boston: Allyn and Bacon.
- Akdeniz Üniversitesi. (Eylül 2012). *Akdeniz Üniversitesi Ön Lisans ve Lisans Eğitim-Öğretim ve Sınav Yönetmeliği*. Erişim: 22 Nisan 2013, <http://oidb.akdeniz.edu.tr/duyurular>
- Aktan, D. Ç. & Çepni, Z. (2010). Ölçme ve değerlendirme dersi kapsamı ve gereklilikleri hakkındaki uzman ve öğretmen görüşleri: pilot çalışma. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 85-92.
- Alsmadi, A. A. (2007). A comparative study of two standard setting techniques. *Social Behaviour and Personality*, 35 (4), 479-486.
- Ankara Üniversitesi. (Şubat 2013). *Ankara Üniversitesi Ön Lisans ve Lisans Eğitim-Öğretim Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.resmigazete.gov.tr>
- Arrasmith, D. G. (1986). *Investigation of judges' errors in angoff and contrasting groups cut of score methods*. Doktora Tezi (University of Massachusetts, Eğitim Bilimleri).
- Atatürk Üniversitesi. (Mayıs 2012). *Atatürk Üniversitesi Ön Lisans ve Lisans Eğitim-Öğretim ve Sınav Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.resmigazete.gov.tr>
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem Yayıncılık.
- Behuniak, P. & Tucker, C. (1992). The potential of criterion-referenced tests with projected norms. *Applied Measurement in Education*, 5(4), 337-353.
- Boğaziçi Üniversitesi. (Ağustos 2011). *Boğaziçi Üniversitesi Lisans Eğitim ve Öğretim Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.resmigazete.gov.tr>
- Calmorin, L. P. & Calmorin, M. A. (2007). *Research methods and thesis writing*. Manila: Rex Book Store.

- Coaley, K. (2010). *An introduction to psychological assessment and psychometrics*. London: Sage Publications.
- Cohen, R. J. & Swerdlik, M. E. (2009). *Psychological testing and assessment: An introduction to tests and measurement*. Boston: McGraw Hill.
- Cizek, G. J. & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Thousands Oaks, CA: Sage Publications.
- Clark-Carter, D. (2005). *Quantitative psychological research: a student handbook*. Hove ve New York: Psychology Press.
- Çetin, S. (2011). *İşaretleme ve angoff standart belirleme yöntemlerinin karşılaştırılması*. Doktora Tezi (Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü).
- Deliceoğlu, G. (2009). *Futbol yetilerine ilişkin dereceleme ölçeğinin genellenebilirlik ve klasik test kuramına dayalı güvenilirliklerinin karşılaştırılması*. Doktora Tezi (Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü).
- Dicle Üniversitesi. (Ağustos 2012). *Dicle Üniversitesi Eğitim-Öğretim ve Sınav Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.dicle.edu.tr/birim/ogrisler/Mevzuat/index.htm>
- Dokuz Eylül Üniversitesi. (Ağustos 2012). *Dokuz Eylül Üniversitesi Ön Lisans ve Lisans Öğretim ve Sınav Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.resmigazete.gov.tr>
- Domino, G. & Domino, M. L. (2006). *Psychological testing: An introduction*. Cambridge: Cambridge University Press.
- Ege Üniversitesi. (Ağustos 2009). *Ege Üniversitesi Eğitim ve Öğretim Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.resmigazete.gov.tr/eskiler/2009/08/20090812-4.htm>
- Erciyes Üniversitesi. (Temmuz 2012). *Erciyes Üniversitesi Ön Lisans ve Lisans Eğitim-Öğretim Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.erciyes.edu.tr>
- Erkuş, A. (2004). The proposal of a new conceptualization for validity and criterion-referenced assessment. *Eurasian Journal of Educational Research*, 16, 113-117.
- Freeman, L. & Miller, A. (2001). Norm-referenced, criterion-referenced, and dynamic assessment: What exactly is the point? *Educational Psychology in Practice*, 17(1), 3-16.

- Gelbal, S. & Çetin, S. (2010). Farklı standart belirleme yöntemlerinin geçme puanları üzerine etkisi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 43 (1), 79–95.
- Gündeğer, C. (2012). *Angoff, yes/no ve ebel standart belirleme yöntemlerinin karşılaştırılması*. Yüksek Lisans Tezi. (Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü).
- Gündeğer, C. & Doğan, N. (2012). *Standart belirleme yöntemleriyle elde edilen geçme puanlarının, meb`e ait geçme puanı ile ilişkisinin incelenmesi*. 11. Ulusal Sınıf Öğretmenliği Eğitimi Sempozyumu'nda sunulan bildiri. Rize Üniversitesi Eğitim Fakültesi, Rize, Türkiye, 24-26 Mayıs.
- Hacettepe Üniversitesi. (Eylül 2012). *Hacettepe Üniversitesi Ön Lisans ve Lisans Eğitim-Öğretim ve Sınav Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.eczacilik.hacettepe.edu.tr>
- Jacobson, R. Y. (2008). *Examination of the potential of selected norm-referenced tests and selected locally developed criterion-referenced tests to classify students into performance categories*. Doktora Tezi (University of Nebraska, Eğitim Yönetimi).
- Jekel, J. F. (2007). *Epidemiology, biostatistics and preventive medicine*. Philadelphia: Saunders/Elsevier.
- Johnson, D. L. & Martin, S. (1980). Criterion-referenced testing: New wine in old bottles. *Academic Therapy*, 16(2), 167 - 173.
- Kane, M. (1998). Choosing between examinee centered and test centered standard setting methods. *Educational Assessment*, 5 (3), 129 – 145.
- Kline, P. (2000). *Handbook of psychological testing*. London and New York: Routledge.
- Livingston, S. A. & Zieky, M. J. (1989). A comparative study of standard - setting methods. *Applied Measurement in Education*, 2, 121 – 141.
- MacTurk, R. H. & Neisworth, J. T. (1978). Norm referenced and criterion based measures with preschoolers. *Exceptional Children*, 45(1), 34-39.
- Margolis, M. J. (2011). *The impact of examinee performance information on judges' cut scores in modified-angoff standard setting exercises*. Doktora Tezi (Temple University, Eğitim Psikolojisi).

- Marmara Üniversitesi. (Eylül 2012). *Marmara Üniversitesi Ön Lisans ve Lisans Eğitim-Öğretim ve Sınav Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.marmara.edu.tr>
- Marshall, J. C. & Hales, L. W. (1971). *Classroom test construction*. London: Addison-Wesley Publishing Company.
- McCauley, R. J. (1996). Familiar strangers: Criterion-referenced measures in communication disorders. *Language, Speech, and Hearing Services in Schools*, 27, 122-130.
- McCauley, R. J. & Swisher, L. (1984). Use and misuse of norm-referenced tests in clinical assessment: A hypothetical case. *Journal of Speech and Hearing Disorders*, 49, 338 - 348.
- Mehrens, W. A. & Lehmann, I.J. (1991). *Measurement and evaluation in education and psychology*. Belmont, CA: Wadsworth/Thomson Learning.
- Milli Eğitim Bakanlığı. (Ağustos 2003). *Milli eğitim bakanlığı ilköğretim kurumları yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.mevzuat.gov.tr>
- Milli Eğitim Bakanlığı. (Aralık 2004). *Milli eğitim bakanlığı ortaöğretim kurumları sınıf geçme ve sınav yönetmeliği*. Erişim: 22 Nisan 2013, <http://mevzuat.meb.gov.tr>
- Mohr, A. K. (2006). *Criterion referenced and norm referenced predictors of student achievement: Teacher perceptions of, and correlations between, iowa test of basic skills and the palmetto achievement challenge test*. Doktora Tezi (University of South Carolina, Eğitim Bilimleri).
- Montgomery, P. C. & Connolly, B. H. (1987). Norm-referenced and criterion referenced tests: Use in pediatrics and application to task analysis of motor skill. *Physical Therapy*, 67(12), 1873-1876.
- Muijs, D. (2004). *Doing quantitative research in education*. London: Sage Publications.
- Murphy, K. R. & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications*. New Jersey: Pearson.
- Nartgün, Z. (2007). Aynı puanlar üzerinden yapılan mutlak ve bağıl değerlendirme uygulamalarının notlarda farklılık oluşturup oluşturmadığına ilişkin bir inceleme. *Ege Eğitim Dergisi*, 8(1), 19-40.

- Norcini, J. J. (2003). Setting Standards on Educational Tests. *Medical Education*, 37 (1), 464 – 469.
- Oescher, J., Kirby, P. C. & Paradise, L. V. (1992). Validating state-mandating criterion-referenced achievement tests with norm-referenced test results for elementary and secondary students. *Journal of Experimental Education*, 60(2), 141-150.
- Özgüven, İ. E. (2011). *Psikolojik testler*. Ankara: PDREM.
- Pamukkale Üniversitesi. (Temmuz 2009). *Pamukkale Üniversitesi Lisans Eğitim ve Öğretim Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.resmigazete.gov.tr>
- Paul, R. (1995). *Language disorders from infancy through adolescents: Assessment and interventions*. St. Louis: Mosby – Year Book Inc.
- Pester, A. M. (2003). *Language intervention effects of norm-referenced and criterion referenced test scores*. Yüksek Lisans Tezi (Miami University, Konuşma Patolojisi).
- Sakarya Üniversitesi. (Ağustos 2011). *Sakarya Üniversitesi Lisans ve Ön Lisans Eğitim-Öğretim ve Sınav Yönetmeliği*. Erişim: 22 Nisan 2013, <http://www.resmigazete.gov.tr>
- Shin, S. K. (2004). *How much is enough?: Setting and validating performance standards and cut scores for the web-based English as a second language placement exam at UCLA*. Doktora Tezi (University of California, Uygulamalı Dil Bilimleri).
- Smith, T. N. (2011). *Using stratified item selection to reduce the number of items rated in standard setting*. Doktora Tezi (University of South Florida, Psikoloji Bölümü).
- Sönmez, V. & Alacapınar, F. G. (2011). *Örneklendirilmiş bilimsel araştırma yöntemleri*. Ankara: Anı Yayıncılık.
- Tanrıverdi, S. (2006). *Standart belirleme yöntemlerinin geçme puanları üzerine etkisi*. Yüksek Lisans Tezi (Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü).
- Taşdelen, G. (2009). *Nedelsky ve angoff standart belirleme yöntemlerinin genellenabilirlik kuramı ile karşılaştırılmasına ilişkin bir araştırma*. Yüksek Lisans Tezi (Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü).
- Toprakçı, E., Baydemir, G., Koçak, A. & Akkuş, Ö. (2007). *Eğitim fakültelerinin eğitim-öğretim ve sınav yönetmeliklerinin karşılaştırılması*. 16. Ulusal Eğitim Bilimleri

Kongresi'nde sunulan bildiri. Gaziosmanpaşa Üniversitesi Eğitim Fakültesi, Tokat, Türkiye, 05-07 Eylül.

Turgut, M. F. & Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem Yayıncılık.

Urbina, S. (2004). *Essentials of psychological testing*. New York: Wiley

Visintainer, C. (2002). *The relationship between two state-mandated, standardized tests using norm-referenced terranova and the criteria-referenced, performance assessment developed for the maryland school performance assessment program*. Doktora Tezi (Wilmington College, Innovation and Leadership).

Yıldırım, C. (2011). *Bilim felsefesi*. İstanbul: Remzi Kitapevi.

EKLER

EK 1: Çalışmada Kullanılan Ölçme ve Değerlendirme Başarı Testinin Soruları

Aşağıda yer alan ölçme ve değerlendirme dersine ilişkin başarı testi, yüksek lisans çalışması kapsamında hazırlanmıştır. Toplam 20 soru yer almaktadır ve cevaplarınızı kitapçık üzerine işaretleyebilirsiniz. Sonuçlar toplu halde değerlendirilecektir ve bireysel sonuçlarınız gizli tutulacaktır. Çalışmaya katkılarınızdan dolayı teşekkür ederiz.

Mutlaka Ad, Soyad ve Bölümünüzü belirtiniz.

Ad :

Soyad :

Bölüm:

Ölçme ve Değerlendirme Dersi Dönem Sonu Notunuz:

1. Aşağıdaki örneklerin hangisinde sözü edilen ölçme işleminde bağıl (keyfi, itibari) sıfır söz konusu değildir?

- A) Ankara ili Çankaya ilçesinin deniz seviyesinden yüksekliği 870 metredir.
- B) Öğrencilerin felsefe dersine karşı tutum puanları ortalaması 207'dir.
- C) Tarih sorularından en az dört tanesine doğru cevap veremeyenler sıfır puan almıştır.
- D) Bugün hava sıcaklığı 15 °C'dir.
- E) Bu kitabın ilk 4 sayfasında kitabın kapsamı tanıtılmaktadır.

2. Aşağıdaki durumların hangisinde kriter referanslı (ölçüt dayanaklı) değerlendirme, norm referanslı (norm dayanaklı) değerlendirmeden daha uygun olur?

- A) Öğrencilerin mezun edilip edilmemelerine karar verilmesinde
- B) Kontenjanları sınırlı olan yükseköğretim programlarına öğrenci seçiminde
- C) Öğrencilerin hazırladıkları fen bilimleri projelerinden en iyisinin belirlenmesinde
- D) Okulda dereceye girecek öğrencilerin belirlenmesinde
- E) Bir sınıf içindeki en başarısız öğrencilerin belirlenmesinde

3. Pazarlama elemanları seçiminde kullanılmak üzere bir test geliştirilmiş ve işe alımlarda bu test uygulanmıştır. Daha sonra, işe alınan elemanların ilk aydaki başarılı satış sayıları ile işe girişte aldıkları test puanları arasındaki uyum incelenmiştir.

Yukarıda belirtilen inceleme, uygulanan testin hangi özelliği hakkında bilgi verir?

- A) Yordama geçerliği
- B) Yapı geçerliği
- C) Kapsam geçerliği
- D) İç tutarlığı
- E) Puanlayıcılar arası güvenilirliği

4. Yazılı yoklamalardaki sorularda, öğrencilerin cevaplarını düşünerek bulmaları ve buldukları bilgileri organize ederek yazmaları gerekir. Bu tür sorularla üst düzeydeki zihinsel beceriler ölçülebilir.

Aşağıdaki sorulardan hangisi, yazılı yoklamaların bu avantajlarını ortaya koyan uygun bir örnek değildir?

- A) Enflasyonu düşürmek için ne tür önlemler alınabilir? Örnek vererek irdelleyiniz.
- B) Avrupa Birliği'ne katılma konusundaki görüşleriniz nelerdir? Bu şekilde düşünmenizin nedenlerini açıklayınız.
- C) İzlediğiniz en iyi sinema filmi hangisidir? Bu filmi neden beğendiğinizi gerekçeleriyle açıklayınız.
- D) Başarı ile yetenek arasındaki ilişkiyi bir örnekle açıklayınız. Sizce insan yaşamında hangisi daha ön plandadır?
- E) Ölçme nasıl tanımlanır? Eğitimde yapılan ölçmeler kaç gruba ayrılır?

5. Sosyal bilgiler sınavında sorulan bir soru şöyledir:

Türkiyenin pek çok sınır komşusu vardır. En uzun sınırı paylaştığımız sınır komşumuz aşağıdakilerden hangisidir?

- A) Bulgaristan'la sınırimiz en uzundur.*
- B) Irak'la sınırimiz en uzundur.*
- C) İran'la sınırimiz en uzundur.*
- D) Suriye'yle sınırimiz en uzundur.*
- E) Yunanistan'la sınırimiz en uzundur.*

Yukarıdaki sorunun kusurları arasında aşağıdakilerden hangisi yoktur?

- A) Kökte yazım hatasının olması
- B) Kökte gereksiz sözcük kullanılması
- C) Kök ile seçenekler arasında ifade uyumsuzluğu olması
- D) Seçeneklerde aynı sözcüklerin tekrarlanması
- E) Seçeneklerin uygun olmayan bir sırada verilmesi

6. Bir ölçme aracının özelliklerinden bazıları şunlardır:

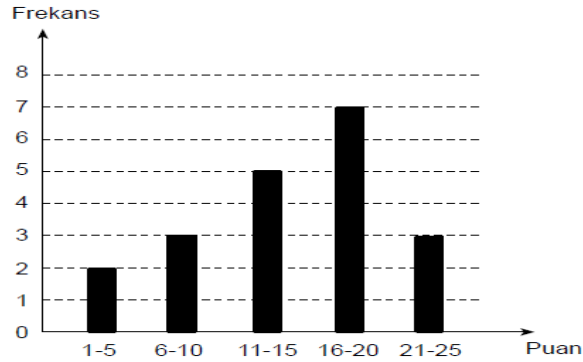
- Bir işin yapılması sırasında nelerin, hangi sırada ve nasıl yapılacağını göstermesi nedeniyle öğretim amaçlı da kullanılabilir.
- Gözlenecek davranışların varlığını veya yokluğunu gösterme olanağı tanır.
- Performansı oluşturan davranışlardan ne kadarına sahip olduğunu belirlemeyi kolaylaştırır.
- Performansın en önemli ve gözlenebilir yanlarını kapsar.
- Performansı oluşturan davranışların her birine ne düzeyde sahip olduğu hakkında bilgi veremez.

Bu ölçme aracı aşağıdakilerden hangisidir?

- A) Dereceleme ölçeği
- B) Kontrol listesi
- C) Tutum ölçeği
- D) Anket
- E) Puanlama yönergesi (rubrik)

7. – 8. SORULARI AŞAĞIDAKİ BİLGİLERE GÖRE CEVAPLAYINIZ.

5A sınıfındaki öğrencilerin 25 soruluk matematik testinden aldıkları puanlar gruplanarak aşağıdaki frekans grafiğinde gösterilmiştir.



7. 5A sınıfının matematik testi puanlarının aritmetik ortalaması kaçtır?

- A) 5,7 B) 7,4 C) 14,5 D) 18,6 E) 20

8. 5A sınıfının matematik testi puanlarının tepe değeri (mod) kaçtır?

- A) 5 B) 7 C) 8 D) 18 E) 25

9. Durum değerlendirme ya da düzey belirleme denen ölçme ve buna bağlı değerlendirme sürecinde temel amaç aşağıdakilerden hangisidir?

- A) Öğretim programında belirtilen kazanımların gerçekleşme derecesini belirlemek
 B) Öğrencilerin yaşadıkları öğrenme güçlüklerini ortaya çıkarmak
 C) Öğrencilerin sosyal, psikolojik ve ekonomik durumlarını belirlemek
 D) Öğrencileri belli özelliklere göre sıralayıp ilgili programlara yerleştirmek
 E) Öğrencilerin kavramsal gelişim düzeylerini ve kavram yanlışlarını tespit etmek

10. Aşağıdaki değerlendirme süreçlerinin hangisinde bağıl ölçüt kullanılmıştır?

- A) Mezun olmak için diploma notunun 4 üzerinden en az 2 olması
 B) Ortalamanın 10 puan üzerinde alanların sınıfı geçmesi
 C) Askere ağırlığı 150 kg'ın altındaki erkeklerin alınması
 D) Soruların %80'ini doğru cevaplayanların bilgi yarışması elemelerine katılması
 E) Yüksek lisansa girebilmek için yabancı dil sınavından en az 70 puan alınması

11. Matematik dersinde bilgisayar programlarından yararlanmanın öğrencilerin matematik başarısına etkisini ölçmeyi planlayan bir öğretmen, bir sınıfta geleneksel öğretim yöntemini kullanırken diğer bir sınıfta ise işlenen her konudan sonra bilgisayar programıyla alıştırmalar yaptırmaktadır.

Bu araştırmadaki bağımsız değişken aşağıdakilerden hangisidir?

- A) Öğretmen özellikleri
- B) Matematik notları
- C) Öğretim yöntemleri
- D) Sınıf mevcutları
- E) Öğrenci nitelikleri

12. Aynı öğrencilere uygulanan ve bağıl değerlendirme yapılan iki sınavla ilgili aritmetik ortalama ve standart sapma değerleri aşağıda verilmiştir.

Sınav	Aritmetik ortalama	Standart sapma	Ali'nin puanı	Ayşe'nin puanı
I	55	15	70	?
II	60	5	?	68

Buna göre, aşağıdaki yorumlardan hangisi kesinlikle doğrudur?

- A) Ali birinci sınavdaki soruların hepsini doğru yanıtlamıştır.
- B) Ali ikinci sınavdan da 70 alırsa ilk sınava göre daha başarısız olmuş olur.
- C) Birinci ve ikinci sınav sonuçları arasındaki korelasyon yüksektir.
- D) Ayşe'nin ikinci sınavdaki başarısı Ali'nin ilk sınavındaki başarısından daha iyidir.
- E) İkinci sınavda puanların ranjı ilk sınava göre daha büyüktür.

13. Yabancı dil sınavında Duru 80, Berke 40 puan almıştır.

Duru'nun yabancı dil bilgisinin Berke'nin yabancı dil bilgisinin iki katı kadar olduğunu öne süren bir kişi, bu ölçme sonuçlarının aşağıdaki ölçek türlerinin hangisiyle elde edildiğini kabul etmektedir?

- A) Eşit aralıklı
- B) Sıralama
- C) Dereceleme
- D) Sınıflama
- E) Eşit oranlı

14. I. İş deneyimi beş yılın üstünde olanlar işe alınmıştır.
 II. Zeynep, tatile gitmeden önce tartılmış; 90 kg gelmiştir.
 III. Merve, ortalamanın üstünde puan alarak sınıfını geçmiştir.
 IV. Elif'in ALES puanı 80'dir.

Yukarıdakilerde hangilerinde “Değerlendirme” işlemi yapılmıştır?

- A) I ve II B) I ve III C) II ve III D) II ve IV E) III ve IV

15. Bir Kuzey Avrupa ülkesinde hemşirelik lisans programı mezunlarının çalışmaya başlayabilmeleri için yeterli düzeyde bilgi ve beceri sahibi olduklarını çoktan seçmeli bir testle göstermeleri gerekmektedir. Bu testte adayların başarılı sayılmaları için doğru cevaplamaları gereken soru sayıları meslek standartları doğrultusunda belirlenir ve yeterlik düzeyi için bir kesme puanı oluşturulur. Testten başarılı olan adaylar, çalışmaya başlamak için gerekli olan sertifikalarını almaya hak kazanırlar.

Yapılan değerlendirmenin amacına ve ölçütüne göre türü aşağıdakilerin hangisinde doğru olarak verilmiştir?

<u>Amacına göre</u>	<u>Ölçütüne göre</u>
A) Biçimlendirme ve yetiştirme	Mutlak
B) Tanıma ve yerleştirme	Bağlı
C) Değer biçme	Bağlı
D) Tanıma ve yerleştirme	Mutlak
E) Değer biçme	Mutlak

16. Bir öğretmen, kelime bilgisini yoklayan bir test hazırlamış, bunu birkaç hafta arayla öğrencilere iki kez uygulamış ve bu iki uygulamadan elde edilen puanlar arasındaki korelasyonu 0,90 olarak hesaplamıştır.

Hesaplanan bu korelasyon değerine dayanarak aşağıdakilerden hangisi söylenebilir?

- A) Sınıfın %10'u başarısızdır.
- B) Test puanlarının kararlılığı yüksektir.
- C) Test puanlarında %10 hata vardır.
- D) Sınıfın %10'u ortalamanın altında puan almıştır.
- E) Test, öğrenme düzeyi farklarının %90'ını açıklamaktadır.

17. Öğretmenler genellikle söz dinleyen ve kurallara uygun davranan öğrencilere sempati duyarlar. Yapılan araştırmalar, öğretmenlerin tepkilerinin, öğrencilerin derslerdeki başarısızlıklarından çok, kurallara uymamalarına yönelik olduğunu ortaya koymaktadır. Öğrencilere bu konuda bir açıklama yapılmasa da onlar, kurallara uygun davrandıklarında bunun başarı olasılığını etkileyebileceğini, sınavda doğru yanıtlar vermeseler de iyi not alabileceklerini düşünürler. Bu durumda, kurallara karşı çıktıklarında cezalandırılacaklarını ya da öğretmenleri tarafından dışlanabileceklerini öğrenirler.

Kurallara uyan öğrencilere daha yüksek puan vererek gerçekleştirilen bu ölçmelerde aşağıdakilerden hangisi yapılmış olur?

- A) Rastlantısal hata
- B) Gözlem yetersizliği
- C) Sabit hata
- D) Mantık hatası
- E) Sistemik hata

18. – 20. SORULARI AŞAĞIDAKİ BİLGİLERE GÖRE CEVAPLAYINIZ.

Aşağıdaki tabloda 10 soruya ait madde gücü (p) ve ayırtedicilik (r) değerleri verilmiştir.

Soru	1	2	3	4	5	6	7	8	9	10
p	0,35	0,05	0,15	0,90	0,70	0,40	0,15	0,75	0,15	0,95
r	0,25	0,65	0,15	0,05	-0,35	0,40	0,10	0,60	-0,40	0,25

18. Uygulama sonucunda yukarıdaki değerleri elde eden bir öğretmen, ayırtediciliği düşük olan maddeleri düzeltilemediği için testten çıkarmak istemektedir.

Bu maddelerin çıkarılması aşağıdakilerden hangisine neden olur?

- A) Testin ölçüt geçerliğinin düşmesine
- B) Kalan maddelerin güçlük indeksinin düşmesine
- C) Çeldiricilerin iyi çalışmamasına
- D) Testin iç tutarlığının düşmesine
- E) Testin kapsam geçerliğinin düşmesine

19. Tablodaki maddelerden hangisi hem kolay hem de ayırtedici bir sorudur?

- A) 3
- B) 4
- C) 8
- D) 9
- E) 10

20. Tablodaki maddelerden hangisi, testle ölçülen özellikler açısından oldukça üst düzey yeterliğe sahip olan öğrencileri seçmek için kullanılabilir?

- A) 1
- B) 2
- C) 5
- D) 7
- E) 10

TEST SONA ERMİŞTİR.

EK 2: Çalışmada Kullanılan Ölçme ve Değerlendirme Başarı Testinin Cevapları

Cevaplar	
1 E	11 C
2 A	12 D
3 A	13 E
4 E	14 B
5 E	15 E
6 B	16 B
7 C	17 E
8 D	18 E
9 A	19 C
10 B	20 B

EK 3: ITEMAN Sonuçlarına Göre Veri Toplama Aracına Seçilen Soruların Madde Güçlük ve Madde Ayırt Edicilik Gücü İndeksleri

MADDE İSTATİSTİKLERİ		
Soru No	Madde Ayırt Edicilik Gücü İndeksi (Çift Serili)	Madde Güçlük İndeksi
1	0,69	0,54
2	0,56	0,36
3	0,44	0,57
4	0,49	0,77
5	0,38	0,52
6	0,53	0,38
7	0,80	0,85
8	0,56	0,80
9	0,36	0,50
10	0,66	0,58
11	0,33	0,75
12	0,42	0,15
13	0,68	0,85
14	0,56	0,81
15	0,55	0,11
16	0,75	0,85
17	0,36	0,68
18	0,57	0,54
19	0,69	0,19
20	0,43	0,26

EK 4: Angoff Yöntemi'ne Göre Test Maddelerinin Uzmanlarca Değerlendirilmesi

Uzmanlar	Test Maddeleri																				Toplam
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
U-1	0,8	0,65	0,85	0,5	0,9	0,7	0,9	0,9	0,7	0,85	0,8	0,6	0,85	0,85	0,55	0,8	0,75	0,5	0,95	0,7	15,1
U-2	0,8	0,5	0,9	0,8	0,9	0,8	0,8	0,6	0,7	0,8	0,5	0,3	0,5	0,6	0,4	0,6	0,6	0,5	0,6	0,6	12,8
U-3	0,9	0,4	0,7	0,9	0,9	0,6	0,85	0,9	0,8	0,85	0,6	0,4	0,85	0,7	0,6	0,6	0,8	0,6	0,6	0,6	14,15
U-4	0,8	0,8	0,7	0,85	0,7	0,6	0,9	0,9	0,5	0,6	0,6	0,3	0,5	0,75	0,25	0,8	0,6	0,7	0,8	0,75	13,4
U-5	0,7	0,5	0,5	0,6	0,3	0,3	0,5	0,4	0,3	0,5	0,6	0,7	0,7	0,7	0,4	0,4	0,7	0,3	0,4	0,4	9,9
U-6	0,75	0,7	0,85	0,65	0,85	0,6	0,9	0,9	0,5	0,8	0,75	0,4	0,85	0,8	0,6	0,7	0,4	0,8	0,9	0,8	14,5
U-7	0,6	0,7	0,6	0,45	0,55	0,5	0,6	0,8	0,5	0,6	0,4	0,3	0,5	0,6	0,4	0,4	0,45	0,55	0,6	0,5	10,6
U-8	0,6	0,7	0,7	0,5	0,6	0,4	0,4	0,7	0,4	0,5	0,5	0,4	0,6	0,5	0,4	0,6	0,5	0,6	0,4	0,4	10,4
U-9	0,7	0,65	0,9	0,95	0,95	0,95	1	1	0,95	0,9	0,95	0,6	0,6	0,95	0,8	0,95	0,95	0,7	0,8	0,9	17,15
U-10	0,8	0,7	0,85	0,7	0,75	0,7	0,9	0,9	0,55	0,8	0,85	0,55	0,9	0,75	0,55	0,65	0,75	0,5	0,75	0,75	14,65
U-11	0,8	0,5	0,7	0,85	0,65	0,9	0,6	0,65	0,55	0,85	0,45	0,5	0,7	0,6	0,55	0,45	0,55	0,5	0,6	0,6	12,55
U-12*	0,75	0,8	0,75	0,65	0,6	0,5	0,3	0,3	0,5	0,6	0,8	0,15	0,6	0,6	0,6		0,7	0,35	0,35	0,3	10,2
U-13*	0,6	0,65	0,65	0,7	0,75	0,7	0,7	0,75	0,6	0,7	0,7	0,65	0,65	0,75	0,65	0,7	0,7	0,7	0,75	0,75	13,8
U-14*	0,85	0,8	0,8	0,7	0,6	0,7	0,6	0,6	0,7	0,85	0,9	0,75	0,9	0,75	0,6	0,65	0,65	0,65	0,45	0,45	13,95

* Uzmanlar arası uyumu olumsuz etkilemesi nedeniyle çalışma kapsamı dışında tutulmuştur.

EK 5: Nedelsky Yöntemi'ne Göre Test Maddelerinin Uzmanlarca Değerlendirilmesi

Test Maddeleri	Uzmanlar																											
	U-1		U-2		U-3		U-4		U-5		U-6		U-7		U-8		U-9		U-10		U-11*		U-12*		U-13*		U-14*	
1	2	0,33	3	0,50	3	0,50	1	0,25	2	0,33	2	0,33	2	0,33	2	0,33	2	0,33	2	0,33	2	0,33	1	0,25	2	0,33	2	0,33
2	2	0,33	2	0,33	3	0,50	0	0,20	4	1,00	3	0,50	3	0,50	1	0,25	1	0,25	3	0,50	2	0,33	2	0,33	4	1,00	2	0,33
3	2	0,33	2	0,33	3	0,50	4	1,00	4	1,00	3	0,50	4	1,00	1	0,25	2	0,33	3	0,50	3	0,50	3	0,50	2	0,33	1	0,25
4	4	1,00	4	1,00	4	1,00	4	1,00	4	1,00	4	1,00	4	1,00	4	1,00	1	0,25	3	0,50	2	0,33	1	0,25	4	1,00	2	0,33
5	3	0,50	4	1,00	2	0,33	1	0,25	4	1,00	2	0,33	4	1,00	1	0,25	1	0,25	2	0,33	2	0,33	2	0,33	1	0,25	3	0,50
6	2	0,33	2	0,33	1	0,25	2	0,33	4	1,00	2	0,33	2	0,33	2	0,33	2	0,33	2	0,33	2	0,33	2	0,33	3	0,50	2	0,33
7	2	0,33	4	1,00	2	0,33	0	0,20	4	1,00	2	0,33	0	0,20	3	0,50	1	0,25	3	0,50	3	0,50	4	1,00	2	0,33	2	0,33
8	3	0,50	3	0,50	3	0,50	0	0,20	4	1,00	3	0,50	0	0,20	3	0,50	1	0,25	3	0,50	4	1,00	4	1,00	3	0,50	2	0,33
9	1	0,25	2	0,33	3	0,50	1	0,25	1	0,25	2	0,33	2	0,33	1	0,25	0	0,20	2	0,33	1	0,25	2	0,33	3	0,50	1	0,25
10	3	0,50	3	0,50	2	0,33	0	0,20	4	1,00	3	0,50	2	0,33	2	0,33	1	0,25	2	0,33	3	0,50	1	0,25	4	1,00	1	0,25
11	3	0,50	3	0,50	2	0,33	1	0,25	4	1,00	3	0,50	2	0,33	2	0,33	1	0,25	3	0,50	2	0,33	4	1,00	4	1,00	2	0,33
12	2	0,33	0	0,20	2	0,33	1	0,25	2	0,33	3	0,50	3	0,50	0	0,20	1	0,25	1	0,25	3	0,50	0	0,20	2	0,33	2	0,33
13	0	0,20	2	0,33	3	0,50	1	0,25	3	0,50	3	0,50	2	0,33	1	0,25	1	0,25	3	0,50	3	0,50	4	1,00	4	1,00	1	0,25
14	1	0,25	4	1,00	2	0,33	0	0,20	2	0,33	4	1,00	2	0,33	1	0,25	1	0,25	4	1,00	4	1,00	2	0,33	0	0,20	3	0,50
15	2	0,33	2	0,33	2	0,33	0	0,20	2	0,33	2	0,33	0	0,20	1	0,25	1	0,25	2	0,33	2	0,33	2	0,33	3	0,50	2	0,33
16	2	0,33	2	0,33	2	0,33	0	0,20	4	1,00	2	0,33	1	0,25	1	0,25	1	0,25	2	0,33	3	0,50	4	1,00	3	0,50	2	0,33
17	2	0,33	2	0,33	2	0,33	1	0,25	4	1,00	2	0,33	0	0,20	1	0,25	2	0,33	2	0,33	3	0,50	2	0,33	4	1,00	1	0,25
18	1	0,25	1	0,25	2	0,33	0	0,20	4	1,00	2	0,33	0	0,20	1	0,25	1	0,25	2	0,33	2	0,33	0	0,20	3	0,50	1	0,25
19	2	0,33	2	0,33	3	0,50	0	0,20	4	1,00	3	0,50	0	0,20	3	0,50	1	0,25	2	0,33	4	1,00	1	0,25	3	0,50	1	0,25
20	0	0,20	0	0,20	2	0,33	0	0,20	3	0,50	3	0,50	0	0,20	1	0,25	1	0,25	2	0,33	4	1,00	1	0,25	2	0,33	2	0,33
Toplam		7,48		9,65		8,42		6,08		15,58		9,50		7,98		6,78		5,28		8,42		10,42		9,48		11,62		6,42

* Uzmanlar arası uyumu olumsuz etkilemesi nedeniyle çalışma kapsamı dışında tutulmuştur.

EK 6: İzin Yazıları

T.C.
ERCİYES ÜNİVERSİTESİ REKTÖRLÜĞÜ
Öğrenci İşleri Daire Başkanlığı

07/03/2013

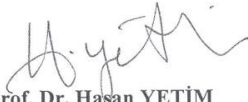
Sayı : 14065294-044/0390 - 0223
Konu : Çalışma İzni

HACETTEPE ÜNİVERSİTESİ REKTÖRLÜĞÜNE

06100-Sıhhiye/ANKARA

İlgi: 22/02/2013 tarihli ve 88600825/240-736 sayılı yazımız

Üniversiteniz Sosyal Bilimler Enstitüsü Eğitim Bilimleri Anabilim Dalı Ölçme ve Değerlendirme Bilim Dalı Yüksek Lisans Programı öğrencisi **İsmail ÇUKADAR**'ın "Norm Dayanlı Değerlendirmelerle Kriter Dayanlı Değerlendirmelerin Karşılaştırılmasına İlişkin Bir Çalışma" konulu tezine ait testini Üniversitemiz Eğitim Fakültesi Eğitim Bilimleri Bölümü Rehberlik ve Psikolojik Danışmanlık Eğitimi Anabilim Dalı 2. Sınıf öğrencilerine uygulama yapma isteği bizzat kendisinin gerçekleştirmesi kaydıyla uygun görülmüştür. Bilgilerinizi arz ederim.


Prof. Dr. Hasan YETİM
Rektör a.
Rektör Yardımcısı

HACETTEPE ÜNİVERSİTESİ
Sosyal Bilimler Enstitüsü
Tarih: 14.03.2013
Sayı: 1810

38039 Melikgazi-KAYSERİ
Tel: 0(352) 437 49 47 Fax: 0(352) 437 20 23

e-mail: ogridbsk@erciyes.edu.tr

Hasan Bey

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı : İsmail ÇUKADAR
Doğum Yeri ve Tarihi : Andırın/K.MARAŞ - 22/02/1985

Eğitim Durumu

Lisans Öğrenimi : Boğaziçi Üniversitesi - Rehberlik ve Psikolojik Danışmanlık
Yüksek Lisans Öğrenimi : Hacettepe Üniversitesi - Eğitimde Ölçme ve Değerlendirme
Bildiği Yabancı Diller : İngilizce (TOEFL: 85)

İş Deneyimi

Çalıştığı Kurumlar : Assessment Sysyems İnsan Kaynakları Danışmanlık Merkezi – (2007/2010)
: MEB Serdar Aksun Ortaokulu – (2011/?)

İletişim

E-Posta Adresi : ismail.cuhadar@gmail.com

Tarih : 16 Temmuz 2013