

T.C.
ABANT İZZET BAYSAL ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

ANGOFF, NEDELSKY VE EBEL STANDART BELİRLEME
YÖNTEMLERİ İLE BELİRLENEN KESME PUANLARININ
KARŞILAŞTIRILMASI

Okan DEMİR

BOLU-2014

T.C.
ABANT İZZET BAYSAL ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

ANGOFF, NEDELSKY VE EBEL STANDART BELİRLEME
YÖNTEMLERİ İLE BELİRLENEN KESME PUANLARININ
KARŞILAŞTIRILMASI

Yüksek Lisans Tezi

Hazırlayan
Okan DEMİR

Danışman
Yrd. Doç. Dr. İ. Alper KÖSE

BOLU-2014

ETİK SÖZLEŞME

Yüksek lisans tezi olarak sunduğum, “Angoff, Nedelsky ve Ebel Standart Belirleme Yöntemleri ile Belirlenen Kesme Puanlarının Karşılaştırılması” başlıklı çalışmanın yazılmasında, bilimsel ve etik kurallara uyulduğunu, başkalarının eserlerinden yararlanılması durumunda atıfta bulunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, tezin tamamının ya da bir kısmının bu üniversite veya başka bir üniversitede bir tez çalışması olarak sunulmadığını beyan ederim. 18.06.2014

OKAN DEMİR

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ'NE,

Okan DEMİR'e ait “Angoff, Nedelsky ve Ebel Standart Belirleme Yöntemleri ile Belirlenen Kesme Puanlarının Karşılaştırılması” adlı çalışma, jürimiz tarafından Ölçme ve Değerlendirme Anabilim Dalında YÜKSEK LİSANS TEZİ olarak kabul edilmiştir **(18.06.2014)**.

Akademik Unvan ve Adı Soyadı

Üye (Tez Danışmanı)	: Yrd. Doç. Dr. İ. Alper KÖSE	İmza
Üye	: Yrd . Doç. Dr. Sevilay KİLMEN	İmza
Üye	: Yrd . Doç. Dr. Recai AKKAYA	İmza

Eğitim Bilimleri Enstitüsünün Onayı

Doç. Dr. Türkan ARGON
Enstitü Müdürü

ABSTRACT

A COMPARISON OF CUTTING POINTS DETERMINED BY ANGOFF, NEDELSKY AND EBEL STANDARD SETTING METHODS

Okan DEMİR

Measurement and Assessment in Education Program

Thesis Advisor: Assist. Prof. Dr. İ. Alper KÖSE

June, 2014 xv + 76 Pages

In this study, it was aimed to compare the cutting points determined by Angoff, Nedelsky and Ebel methods, one of test-based standard setting methods.

In this study, data were collected from two different groups comprising of students and experts. The mathematics achievement test, comprising of 26 questions which are developed by the researcher with the aim of evaluating the objectives of 9th grade mathematics course related to the subjects of sets and equations-inequalities, is applied to 245 students and obtained achievement scores are used as student data of the study. Expert data of the study are collected by the "Expert Opinion Form" developed by the researcher and containing the decisions given by experts about three different standard methods.

The differences between the percentages of students, considered as successful according to cut-off scores determined by each standard methods, have been tested with the test of difference between two percentages. The scores of students were categorized as successful and unsuccessful according to the cut-off scores determined by three methods. The compliance between three methods has been examined with Cohen's Kappa statistics. For each method, the relationship between the minimum passing scores defined by experts has been evaluated with Pearson-Product Moment Correlation

Coefficient and the difference between the averages of the cut-off scores has been evaluated with paired samples t-test.

At the end of this research, it is seen that the highest cut-off score is determined by Angoff method while the lowest cut-off score is determined by the Nedelsky method and also the percentages of students, considered as successful by taking above the cut-off points, differ statistically significant. When the students are classified as successful and unsuccessful according to calculated cut-off scores and the methods are compared as binary, it is seen that there is a high positive compliance between Angoff and Ebel methods and Ebel and Nedelsky methods while there is a positive compliance between Angoff and Nedelsky methods.

It is seen that there is a low relationship between Angoff and Nedelsky methods in terms of the minimum passing scores obtained from the experts whereas the difference between the averages of minimum passing scores is significant. There is a high level significant relationship between Angoff and Ebel methods while the difference between the averages of minimum passing scores of these two methods is not significant. It is found that there is a low level relationship between Nedelsky and Ebel methods, while there is no significant difference between the averages of minimum passing scores of these two methods.

Key Words: Angoff, Nedelsky, Ebel methods, cut-off score, minimum proficiency level.

ÖZET

ANGOFF, NEDELSKY VE EBEL STANDART BELİRLEME YÖNTEMLERİ İLE BELİRLENEN KESME PUANLARININ KARŞILAŞTIRILMASI

Okan DEMİR

Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Tez Danışmanı: Yrd. Doç. Dr. İ. Alper KÖSE

Haziran, 2014 xv + 76 Sayfa

Bu araştırma ile test merkezli standart belirleme yöntemlerinden olan Angoff, Nedelsky ve Ebel yöntemleri ile belirlenen kesme puanlarının karşılaştırılması amaçlanmıştır.

Araştırmada öğrenciler ve uzmanlar olmak üzere iki farklı gruptan veri toplanmıştır. Dokuzuncu sınıf matematik dersi kümeler ve denklemler-eşitsizlikler öğrenme alanına ait kazanımları ölçmek amacıyla araştırmacı tarafından hazırlanan 26 maddeden oluşan matematik başarı testi 245 öğrenciye uygulanmış, elde edilen başarı puanları araştırmanın öğrenci verileri olarak kullanılmıştır. Araştırmada kullanılan uzman verileri, araştırmacı tarafından geliştirilen ve her bir uzmanın üç farklı standart belirleme yöntemi için vermiş oldukları kararları içeren “Uzman Görüşleri Formu” ile toplanmıştır.

Her bir standart belirleme yönteminin belirlediği kesme puanına göre başarılı kabul edilen öğrenci yüzdeleri arasında anlamlı fark olup olmadığı bağımlı iki yüzde arasındaki farkın testi ile sınanmıştır. Yöntemler tarafından belirlenen kesme puanlarına göre öğrenciler başarılı-başarısız şeklinde kategorize edilip, yöntemler arasındaki uyuma Cohen’s Kappa İstatistiği ile bakılmıştır. Her bir yöntem için uzmanlar tarafından belirlenen minimum geçme puanları arasındaki ilişkiye Pearson Momentler

Çarpımı Korelasyon Katsayısı ile minimum geçme puanlarının ortalamaları arasındaki farka ise Bağımlı Gruplar t Testi ile bakılmıştır.

Araştırma sonucunda en yüksek kesme puanının Angoff yöntemi, en düşük kesme puanının Nedelsky yöntemi tarafından belirlendiği ve farklı standart belirleme yöntemlerine göre kesme puanının üzerinde puan alarak başarılı kabul edilen öğrenci yüzdelerinin anlamlı şekilde farklılaştığı görülmüştür. Hesaplanan kesme puanlarına göre öğrenciler başarılı ve başarısız olarak sınıflandırılıp yöntemler ikili olarak karşılaştırıldığında, Angoff ve Ebel yöntemleri ile Ebel ve Nedelsky yöntemleri arasında yüksek derecede pozitif yönde bir uyumun olduğu, Angoff ve Nedelsky yöntemleri arasında ise pozitif yönde makul bir uyumun olduğu görülmüştür.

Angoff ve Nedelsky yöntemleri arasında, uzmanlardan elde edilen minimum geçme puanları bakımından düşük düzeyde bir ilişki olduğu görülürken, minimum geçme puanlarına ait ortalamalar arasındaki fark anlamlı bulunmuştur. Angoff ve Ebel yöntemleri arasında yüksek düzeyde anlamlı bir ilişki olduğu görülürken, bu iki yönteme ait minimum geçme puanlarının ortalamaları arasındaki fark anlamlı bulunmamıştır. Nedelsky ve Ebel yöntemleri arasında düşük düzeyde bir ilişki olduğu görülürken, bu iki yönteme ait minimum geçme puanlarının ortalamaları arasındaki fark anlamlı bulunmamıştır.

Anahtar Sözcükler: Angoff, Nedelsky, Ebel yöntemleri, kesme puanı, minimum yeterlik düzeyi.

Münire ve Gürsel SOMUNCU'ya

TEŞEKKÜR

Çalışmanın tamamlanmasındaki katkılarından dolayı, araştırmalarımın her safhasında görüşleri ile beni yönlendiren, engin tecrübelerinden faydalandığım değerli hocam ve tez danışmanım Sayın Yrd. Doç. Dr. İ. Alper KÖSE'ye,

Kendisinin öğrencisi olmadığım halde bana güvenip, Eğitimde Ölçme ve Değerlendirme alanına beni yönlendiren Sayın Prof. Dr. Mehmet BAHAR'a, alanın temel bilgi ve becerilerini kazandıran, deneyimlerinden yararlandığım hocam Sayın Doç. Dr. Zekeriya NARTGÜN'e,

Bu alanda, ilk istatistik bilgilerini kazandıran değerli hocam Sayın Yrd. Doç. Dr. Sevilay KİLMEN'e ve istatistik bilgisinden yararlandığım abim Yrd. Doç. Dr. Osman DEMİR'e,

Üzerimden maddi ve manevi desteklerini esirgemeyen, her zaman yanımda olduklarını bildiğim her iki ailemde de bulunan tüm fertlere sevgi ve teşekkürlerimi sunuyorum.

Hayatımın en önemli dönemeçlerinde yanımda olan, sevgisini ve ilgisini en yoğun günlerimde dahi esirgemeyen, tez çalışmalarım süresince desteğini daima hissettiğim eşim Naciye SOMUNCU DEMİR'e ve güzel gülümsemesiyle hayatıma eşsiz bir anlam katan oğlum Yusuf Alptekin'e sonsuz teşekkür ediyorum.

Okan DEMİR

İÇİNDEKİLER

ETİK SÖZLEŞME	iii
ABSTRACT.....	v
ÖZET	vii
İTHAF.....	ix
TEŞEKKÜR	x
İÇİNDEKİLER	xi
TABLolar DİZİNİ.....	xiv
 BÖLÜM I.....	 1
1. GİRİŞ	1
1.1. Problem Durumu.....	1
1.2. Problem Cümlesi.....	4
1.2.1. Alt problemler	4
1.3. Araştırmanın Amacı.....	5
1.4. Araştırmanın Önemi.....	5
1.5. Varsayımlar	6
1.6. Sınırlılıklar	6
1.7. Tanımlar ve Kısaltmalar	6
 BÖLÜM II	 7
2. KURAMSAL TEMELLER VE İLGİLİ LİTERATÜR.....	7

2.1. Standart Belirleme.....	7
2.1.1. Minimum yeterlik düzeyi	10
2.1.2. Kesme puanı	11
2.2. Standart Belirleme Süreci	11
2.3. Standart Belirleme Yöntemleri	13
2.3.1. Test maddelerine dayanan yöntemler.....	13
2.3.2. Testi alanların performansına dayanan yöntemler	20
2.4. İlgili Literatür	22
2.4.1. Yurt içinde yapılan standart belirleme yöntemleri ile ilgili çalışmalar	22
2.4.2. Yurt dışında yapılan standart belirleme yöntemleri ile ilgili çalışmalar	25
BÖLÜM III.....	31
3. YÖNTEM	31
3.1. Araştırmanın Modeli	31
3.2. Verilerin Elde Edildiği Grup.....	31
3.3. Verilerin Toplanması ve Veri Toplama Araçları	33
3.4. Verilerin Analizi	38
BÖLÜM IV	41
4. BULGULAR VE YORUM.....	41
4.1. Birinci Alt Probleme Ait Bulgular ve Yorumlar.....	41
4.2. İkinci Alt Probleme Ait Bulgular ve Yorumlar	46
4.3. Üçüncü Alt Probleme Ait Bulgular ve Yorumlar	48

BÖLÜM V	52
5. SONUÇLAR VE ÖNERİLER	52
5.1. Sonuçlar	52
5.2. Öneriler	53
5.2.1. Araştırma sonuçlarından çıkarılan öneriler	53
5.2.1. Araştırmacılara yönelik öneriler	54
KAYNAKÇA.....	55
EKLER	63
Ek.1. Matematik Başarı Testi	63
Ek.2. Uzman Görüşleri Formu	68
Ek.3. Angoff yöntemi ile uzmanlardan alınan kararlar	74
Ek.4. Nedelsky yöntemi ile uzmanlardan alınan kararlar	75
Ek.5. Ebel yöntemi ile uzmanlardan alınan kararlar	76

TABLOLAR DİZİNİ

Tablo 2-1. Farklı beş performans düzeyi için kesme puanı örneği	10
Tablo 2-2. Angoff yöntemi ile kesme puanı belirleme örneği	14
Tablo 2-3. Nedelsky yöntemi ile kesme puanı belirleme örneği	16
Tablo 2-4. Ebel yöntemi ile “zorlu” ve “uygunluk” boyutlarının tespit edilmesini gösteren bir örnek	17
Tablo 2-5. Ebel yöntemi ile kesme puanı belirleme örneği	18
Tablo 3.1. Araştırmaya katılan okullar ve okulların öğrenci sayıları	32
Tablo 3-2. Deneme uygulaması sonucu oluşan madde istatistikleri	34
Tablo 3-3. Nihai testin kestirilen istatistikleri	35
Tablo 3-4. Nihai testin uygulanması sonucu oluşan test istatistikleri	36
Tablo 3-5. Bağımlı gruplarda iki oran için genel tablo	38
Tablo 4-1. Yöntemler için bulunan MGP’ler ve kesme puanları	41
Tablo 4-2. Farklı yöntemlerle belirlenen kesme puanları ve bu kesme puanlarına göre başarılı kabul edilen öğrenci sayısı ve yüzdeleri	43
Tablo 4-3. Angoff ve Nedelsky yöntemlerine göre başarılı-başarısız öğrenci sayısını gösteren çapraz tablo	44
Tablo 4-4. Angoff ve Ebel yöntemlerine göre başarılı-başarısız öğrenci sayısını gösteren çapraz tablo	44
Tablo 4-5. Nedelsky ve Ebel yöntemlerine göre başarılı-başarısız öğrenci sayısını gösteren çapraz tablo	45
Tablo 4-6. Yöntemlerin belirlediği kesme puanları, başarılı öğrenci sayı-yüzdesi	

ve z değerleri.....	45
Tablo 4-7. Yöntemler arasındaki uyumu gösteren kappa katsayıları	47
Tablo 4-8. MGP'lere ait betimsel istatistikler ve normallik testi sonuçları	48
Tablo 4-9. Uzmanlardan elde edilen MGP'ler arası pearson momentler çarpımı korelasyon katsayısı sonuçları	49
Tablo 4-10. MGP'lere ait ortalamalar arası farkın anlamlılığı için yapılan t testi sonuçları	50

BÖLÜM I

1. Giriş

Çalışmanın bu bölümünde araştırmanın problem durumu, araştırmanın amacı, araştırmanın önemi, araştırmanın varsayımları ve sınırlılıkları ele alınmıştır.

1.1. Problem Durumu

Hızla değişen ve gelişen dünya, nüfus artışı ve yoğun rekabet nedeniyle günümüzde her sektörde daha nitelikli insan tipi aranır hale gelmiştir. Her türlü işe alımlarda kişide olması beklenen niteliklerin eşik değerinin oldukça yükseldiği görülmektedir. Bu nedenle bireyler, çağın gerektirdiği insan tipinin niteliklerine sahip olabilmek ya da var olan niteliklerini yükseltebilmek adına eğitim süreçlerine katılmak durumunda kalmaktadırlar.

Bireyin davranışlarında kendi yaşantısı yoluyla, kasıtlı ve istendik olarak bir değişme meydana getirme sürecine eğitim denir (Ertürk, 1994). İnsan davranışlarını geliştiren bir sistem olarak görülen eğitimin, her sistemde olduğu gibi girdi, süreç, çıktı ve kontrol gibi öğeleri vardır. Sistemin kontrolü eğitimde değerlendirme ile yapılır. Değerlendirme, eğitim sistemindeki öğelerin koordinasyonunu sağlar, varsa aksayan bir öğe, kendisi de dahil, onun düzeltilmesini sağlar (Baykul, 2000).

Değerlendirme, ölçme sonuçlarının bir ölçütle karşılaştırılarak, bir karara varılması işidir (Turgut, 1993). Ölçütsüz değerlendirme olmaz. Ölçüt, ya önceden, kesin olarak tespit edilmiş bir standarttır ya da gruba bağlı olarak çıkarılmış tipik bir puandır. Ölçüt, değerlendirilecek grup dikkate alınmadan önceden kesin olarak belirtilirse mutlak ölçüt, grubun ortalama başarısı gibi grubun başarısından çıkarılan bir norm ise bağıl ölçüt adını alır. Mutlak ölçüte göre yapılan değerlendirmelere mutlak değerlendirme, bağıl ölçüte göre yapılan değerlendirmelere ise bağıl değerlendirme adı verilir (Livingston ve Zieky, 1982; Tekin, 1979). Ülkemizde her yıl eğitim alanında

yapılan sınavlardan Lisans Yerleştirme Sınavı (LYS), sınava girenleri sıralayarak yerleştirme yapmaktadır. Yine her yıl yapılmaya başlanan Yabancı Dil Sınavı (YDS) belli bir puan üstünde alanları başarılı kabul etmektedir. Bu anlamda LYS’de bağıl değerlendirme, YDS’de mutlak değerlendirme yapıldığı söylenebilir.

Öğrenci başarısının belirlenmesinde, işe alımlarda veya öğrenci seçimlerinde çeşitli testlerden yararlanılmaktadır. Amaca uygun olarak, bu testler bireyin istenilen özelliklere ne kadar sahip olduğunu ortaya çıkaracak şekilde hazırlanmaktadır (Gündeğer, 2012). Pek çok durumda herhangi bir testin sonucunun yorumlanabilmesi, öncelikle bir geçme puanının elde edilmesini zorunlu kılar. Örneğin ünitelere ayrılmış bir konuda, öğrencinin bir üniteden diğerine geçebilmesi için işlenen ünitelerdeki konularda belirli bir bilgi düzeyine ulaşması beklenir. Bu bilgi düzeyi ise geçme ya da kesme puanı ile belirlenir. Birey kesme puanının üzerinde puan almışsa bir sonraki üniteye geçebilir, kesme puanının altında bir puan almışsa da o üniteyi tekrar etmesi beklenir (Taşdelen, 2009). Bazen uygulayıcılar tarafından güvenilir olmayan ölçme sonuçlarına dayanarak karar verildiği, bazen de geçerli olmayan ölçütler seçildiği veya ölçütün yanlış kullanıldığı, hatta bazen hiç ölçüt seçilmeden kararlar verildiği görülmektedir. Bu yanıştan hareketle değerlendirmelerde ölçüt yani standart belirleme işi çok önemlidir (Çetin, 2011).

Standart belirleme, performans düzeylerini ayıran, genellikle kesme ya da geçme puanı olarak adlandırılan ve yeterli düzeyde olan öğrenci ile olmayan öğrenciyi ayırdığı düşünülen performans noktasının belirlenmesi sürecidir (Crocker ve Algina, 1986; Hambleton, 2001; Irwin, 2007). Standart belirleme günümüzde eğitim kurumlarının çıktılarını raporlamada temel amaç haline gelmiştir. Eğitim kurumları ve test merkezleri test sonuçlarına dayanarak karar vermek durumundadırlar (Kubiszyn ve Borich, 2007). Süreç sonunda elde edilen kesme puanı ölçme sonuçlarına bağlı olarak, bireyleri iki veya daha fazla performans düzeyine yerleştirmek amacıyla kullanılır. Elde edilen kesme puanı sınavı alan bireylerin gelecekleri hakkında, bazı testlerde yeterli-yetersiz, geçti-kaldı, bazı testlerde ise başlangıç düzeyde, orta düzeyde, yeterli düzeyde ve ileri düzeyde gibi birden fazla performans kategorisine yerleştirilme durumuna karar vermede kullanılır (Taşdelen, 2009). Bazen de bu kararlar sertifika kararları, yerleştirme kararları, teşvik kararları ve lisanslama kararları şeklinde olabilir. Değerlendirmeler

sonucu verilen bu kararların doğruluğu, ölçme amacıyla kullanılan testin güvenilir ve geçerli olması yanında ölçütün (standartın) doğru belirlenmiş olmasına da bağlıdır (Kubiszyn ve Borich, 2007).

Literatürde bir çok standart belirleme yöntemi vardır ve standart belirlemede hangi yöntemin kullanılacağına bazı faktörler dikkate alınarak karar verilmelidir. Bu faktörlerden bazılarını maliyet, zaman, güçlük, puanlayıcıların sınavı alacak kişileri tanıması, puanlayıcıların sınav içeriği hakkında bilgi sahibi olması, standart belirleme sürecini yapacak kişinin analiz bilgisine sahip olması olarak sıralayabiliriz. Bazı yöntemler vardır ki standart belirleme sürecindeki uygulamaları kısa zaman alırken, yöntemin analizleri bir hayli karmaşıktır. Diğer taraftan öyle yöntemler vardır ki standart belirleme süreci uzun bir süreç gerektirirken, yöntemin analizleri oldukça kolaydır. Bu anlamda hiçbir standart belirleme yöntemi zaman, maliyet ve uygulama kolaylığı öğelerinin bütünüyle diğerlerine karşı avantajlı değildir. Bu sebeple en doğru yöntemi seçme işlemi oldukça zor bir iştir (Taşdelen, 2009).

Farklı standart belirleme yöntemlerinin kesme puanı belirlemede ne tür sonuçlar verdiğini inceleyen bir çok araştırma literatürde mevcuttur (Appleton, 1984; Brennan ve Lockwood, 1980; Chang, 1996; Irwin, 2007; Jaeger, 1989; Tanrıverdi, 2006). Aynı sınava aynı puanlayıcılar tarafından farklı standart belirleme yöntemleri uygulandığında farklı kesme puanlarının ortaya çıktığı, yöntem seçiminin kesme puanları üzerinde etkili olduğu görülmüştür (Irwin, 2007; Kane, 1998). Bu nedenle standart belirleme sürecinde birden fazla standart belirleme yönteminin beraber kullanılması daha doğru sonuçlar verecektir (Cizek, 2001). Kesme puanlarının birbirine yakın sonuçlar vermesi, kesme puanını güçlendirirken; kesme puanlarının birbirinden oldukça farklı sonuçlar vermesi, araştırmacılara gözden geçirilmesi gereken bazı durumların varlığını işaret eder (Irwin, 2007).

Standart belirleme yöntemleri test merkezli ve öğrenci merkezli yöntemler olmak üzere ikiye ayrılır (Jaeger, 1989). Test merkezli yöntemlerde uzmanlar testteki maddeler hakkında, öğrenci merkezli yöntemlerde ise testi alan bireyler hakkında bir yargıya varır. En çok kullanılan test merkezli standart belirleme yöntemleri, Angoff, Nedelsky, Ebel, Yes/No, Ortalama Kestirme Yöntemi, Bookmark, Mapmark iken, en

çok kullanılan öğrenci merkezli standart belirleme yöntemleri ise Sınır Grup Yöntemi ve Karşıt Gruplar Yöntemidir.

Burada bahsedilen standart belirleme yöntemlerinden başka literatürde daha birçok yöntemin varlığından söz edilebilir. Her standart belirleme yönteminin diğerlerine karşı var olan avantajlarının yanında dezavantajları da mevcuttur. Her ne kadar yurt dışında bazı çalışmalar olsa da, yurt içinde eğitimcilerin bu yöntemlerden hangilerini, hangi durumlarda seçecekleri ve nasıl kullanacakları konusunda gerekli bilgiye ulaşabilecekleri ve yöntemlerin birbiriyle ilişkisini gösteren yeterli sayıda çalışmanın olmadığı görülmektedir. Hangi yöntemin daha düşük ya da yüksek kesme puanı belirlediği, hangi yönteme göre başarılı kabul edilen öğrenci sayısının daha az ya da fazla olduğu, hangi yöntemde uzmanlar arasındaki uyumun daha fazla olduğu, hangi yöntemle elde edilen kesme puanının öğrencilerin gerçek performanslarına yakın olduğu bilgisi, sınav uygulayıp bir karar vermek durumunda olan her kurum için bir gerekliliktir. Bu araştırma ile test merkezli standart belirleme yöntemlerinden olan, ayrı ayrı sıkça kullanılmalarına rağmen, aynı anda karşılaştırıldıkları nadir görülen yöntemlerden Angoff, Nedelsky ve Ebel standart belirleme yöntemlerinin belirlediği kesme puanları karşılaştırılıp, yöntemler arasındaki ilişkiler tespit edilmeye çalışılmıştır.

1.2. Problem Cümlesi

Angoff, Nedelsky ve Ebel standart belirleme yöntemleriyle belirlenen kesme puanları arasında nasıl bir ilişki vardır?

1.2.1 Alt problemler

Yukarıda verilen problem cümlesinden hareketle aşağıdaki alt problemlere yanıtlar aranmıştır.

- Üç farklı standart belirleme yöntemine göre belirlenen kesme puanlarının üzerinde puan alarak, başarılı kabul edilen öğrenci yüzdeleri arasında anlamlı bir farklılık var mıdır?
- Yöntemler arasında öğrencilerin başarılı-başarısız olarak sınıflandırılma durumları bakımından uyum var mıdır?

- Yöntemler arasında uzmanlardan elde edilen minimum geçme puanları (MGP) bakımından uyum var mıdır?

1.3. Araştırmanın Amacı

Bu araştırma ile Angoff, Nedelsky ve Ebel gibi farklı üç standart belirleme yöntemi incelenmiş, standart belirleme yöntemlerinin belirlediği kesme puanları karşılaştırılmış, bu puanlara göre başarılı ve başarısız kabul edilen öğrenci yüzdeleri arasındaki fark, yöntemler arasındaki uyum ve uzmanlardan elde edilen minimum geçme puanları arasındaki ilişkiler tespit edilmeye çalışılmıştır.

1.4. Araştırmanın Önemi

Ulaşılabilen uluslararası literatür tarandığında, farklı standart belirleme yöntemlerini karşılaştıran bir çok çalışmanın mevcut olduğu görülmektedir (Appleton, 1984; Brennan ve Lockwood, 1980; Chang, 1996; Irwin, 2007; Jaeger, 1989). Ne var ki ulusal literatürde standart belirleme yöntemleri veya bu yöntemlerin karşılaştırılmasına yönelik yeterince çalışmaya rastlanmamıştır (Çetin, 2011; Gündeğer, 2012; Tanrıverdi, 2006; Taşdelen, 2009; Tülübaş, 2009). Ebel yöntemini konu alan çalışma sayısının diğer yöntemlere göre oldukça az olduğu görülmekle birlikte, Angoff, Nedelsky ve Ebel yöntemlerinin üçünün aynı anda karşılaştırıldığı çalışma sayısı ise literatürde oldukça az bulunmaktadır. Standart oluşturmak, öğrenci başarısını değerlendirmede ve öğrenci hakkında karar vermede önemli bir aşamadır. Bu konuda öğrenci başarısını değerlendiren eğitimcilerin gerekli bilgiye sahip olabilecekleri yeterli sayıda kaynağın olmadığı düşünülmektedir. Bu çalışmanın eğitimcilere standart belirleme yöntemlerini tanıma ve onları kullanma anlamında bilgi sağlayacağı ve bu konuda literatüre katkı yapacağı düşünülmektedir.

1.5. Varsayımlar

Öğrencilerin test maddelerine verdikleri cevapların öğrenme düzeyleri bakımından gerçeği yansıttığı düşünülmektedir. Uzmanların kararlarını öğrencilerin öğrenme düzeylerini göz önüne alarak verdikleri sanılmaktadır. Uzmanların “minimum yeterlik düzeyindeki öğrenci” ifadesiyle aynı kişiyi düşündükleri varsayılmıştır.

1.6. Sınırlılıklar

Araştırma geliştirilen 9. sınıf 1. dönem müfredatını kapsayan 26 soruluk matematik başarı testi ile sınırlıdır. Araştırmanın öğrenci verileri Bolu-Merkez’de bulunan 13 farklı lisenin 245 öğrencisinin başarı testi puanları ile sınırlıdır. Araştırma standart belirleme yöntemlerinden Angoff, Nedelsky ve Ebel standart belirleme yöntemleri ile sınırlıdır. Araştırmada toplanan uzman kararları verileri 19 uzmanın verileri ile sınırlıdır.

1.7. Tanımlar ve Kısaltmalar

Bu araştırmada sıkça kullanılan bazı kavramlar için aşağıda açıklanan tanımlar benimsenmiştir.

Kesme puanı (Geçme puanı): İstenilen performans düzeyinde olan öğrenci ile yeterli düzeyde olmayan öğrenciyi ayırt edebilen uygun performans noktası. Öğrencilerin performans ya da başarı düzeyleri arasındaki ayrım.

MGP (Minimum geçme puanı): Bir uzmanın belirlediği, başarılı olabilmek için alınması gereken minimum puanı.

Minimum yeterlik düzeyi: Minimum geçme puanına ulaşabilmek için gerekli yeterlik düzeyi.

Sınır grup: Minimum yeterlik düzeyindeki öğrencilerden oluşan grup.

Uzman: Alanında en az 2 yıl deneyimli matematik öğretmeni.

BÖLÜM II

2. Kuramsal Temeller ve İlgili Literatür

Bu bölümde standart belirleme, kesme puanı, sınır grup, standart belirleme süreci ve standart belirleme yöntemi ile ilgili kuramsal açıklamalar ve yapılan ilgili araştırmalar yer almaktadır.

2.1. Standart Belirleme:

Bireysel farklılıklara bağlı olarak her öğrenci öğrenme süreci sonunda farklı düzeylerde öğrenir. Bazı öğrenciler öğretilenlerin hepsini, bazıları daha azını öğrenirken, bazıları ise hedeflenen kazanımları edinemezler. Her öğrencinin öğrenme düzeyi farklı olduğundan farklı boyutlarda ve düzeylerde performans tanımları yapılmalıdır. Performans düzey tanımlarının yani yeterliklerin tanımlanmadığı ölçme sistemlerinin kullanıldığı durumlarda ülke genelinde öğretimin niteliğini artırmak oldukça zor olacaktır (Berberoğlu, 2009). Birçok gelişmiş ülkede ve özellikle Amerika Birleşik Devletleri'ndeki büyük okulların çoğu kendi standartlarını kendileri belirler, öğrencileri bu standartlara göre değerlendirip temel, yeterli ve ileri düzey gibi performans düzeylerine ayırırlar (Nitko ve Brookhart, 2007; National Research Council, 2001).

Standart terimi “Ne kadarı yeterli” sorusunun cevabı olarak düşünülebilir. Herhangi bir test için bir geçme puanı belirlendiğinde o testteki performans için standart belirlenmiş olur. Standartlar mutlak ve bağıl olabilir (Livinston ve Zieky, 1982). “Soruların en az %55’ini doğru cevaplayanlar geçecek” ifadesi mutlak bir standart oluştururken; “Puanı grup ortalamasının üstünde olanlar geçecek” ifadesi bağıl bir standart oluşturmaktadır (Gündeğer, 2012). Bağıl standartlar bireyler arası karşılaştırmalara bağlı iken, mutlak standartlar belirlenen bir ölçütle ilgilidir. Bağıl bir standardı temsil eden geçme (kesme) puanını belirlemek çok zor değilken, mutlak bir

standartı temsil eden bir geme (kesme) puanını belirlemek hem zor hem nemli hem de uzun bir sretir (Livinston ve Zieky, 1982).

Standart belirleme (standard setting) kavramı, alan yazında test maddelerinin doėru yanıtlanma olasılıklarından hareketle bir testten alınacak puanları, belirlenen kesme puanları ile belli dzeylere ayırma ve bu dzeyleri tanımlama anlamında kullanılmaktadır (Schulz, 2006; Shen, 2001; Zwick, entrk ve Wang, 2001). Cizek (2001)'e gre standart belirleme iki ya da daha fazla performans derecesini ayıran puanların belirlenmesinde izlenen tanımlanmış sistemli kurallar ya da sreler btndr. Crocker ve Algina (1986)'ya gre standart belirleme, kesme puanının elde edilmesidir. Bir programı bitirip diėer programa geebilmek iin adayların ařmaları gereken belli bir kesme puanı (cut-off score) vardır. Bu kesme puanı testi hazırlayanlar tarafından belirlenmektedir. Belirlenen kesme puanı, testi alan kiřinin performans alanındaki yeterliėi ile ilgili bir karara varmayı saėlar. Kesme puanları test puanları cetvelini iki ya da daha fazla alana ayırarak performans kategorileri oluřturur ve kiřileri sınıflandırır.

Standart belirleme son zamanlarda eėitim kurumlarının ıktılarını raporlamada temel ama haline gelmiřtir. Eėitim kurumları test sonularına dayalı karar vermektedirler. lme aralarından elde edilen puanlara dayalı olarak alınan bu kararlar geti-kaldı kararları, sertifika kararları, teřvik kararları, lisanslama kararları, yerleřtirme kararları, seme kararları, tanı kararları ve rehberlik gibi kurumsal kararlar řeklinde olabilmektedir (Irwin, 2007). Yapılan deėerlendirmeler sonucu verilen bu kararların isabetli olması testin gvenilir ve geerli olması yanında, ltn (standart) de doėru belirlenmiş olmasına baėlıdır. Deėerlendirmede kararın doėruluėu hem lmenin, hem de ltn doėru olmasına baėlıdır. Sınavdan bařarılı sayılabilmek iin alınması gereken minimum puan amaca uygun biimde belirlenmemiř ise, lme iřlemi hatalardan arınık bile olsa ltn hatalı belirlenmesine baėlı olarak hatalı kararlar alınması muhtemeldir (zelik, 1992; Turgut, 1993).

Kesme puanlarına dayalı yapılan geti-kaldı ve dzey belirleme tr deėerlendirmeler bireysel olarak sadece ėrencileri deėil, tmden eėitim sistemini de etkileyebilmektedir. rneėin uluslararası dzeyde yapılan PIRLS (Progress in International Reading Literacy Study), PISA (Program for International Student

Assessment) ve TIMSS (Trends in International Mathematics and Science Study) gibi sınavlar ülkeler arası eğitim düzeylerinin ve eğitim sistemlerinin karşılaştırılmasında etkindirler. Bu nedenle çoğu gelişmiş ülke, eyaletlerine ve okullarına kendi akademik yeterlik standartlarını belirleme imkanı vermektedir (Kubiszyn ve Borich, 2007).

Performans düzeyleri, standart belirleme yöntemlerinden bağımsız olarak, yöntem belirlenmeden tanımlanabilir. Standart belirleme sürecinde önce kategori sayısı belirlenir, kategorilerin tanımları yapılır, her kategorinin ne anlama geldiği hakkında genel tanımlamalara gidilir. Düzeyler arasındaki farklılıkları ayırtırmayı zorlaştıracığından dolayı, üç veya dörtten fazla kategori çok tavsiye edilen bir durum değildir (Zieky ve Perie, 2004).

Örneğin performans düzeyleri dört kategoriye ayrılıp, her düzeyin tanımı aşağıdaki gibi yapılabilir (Çetin, 2011):

- İleri düzey: Yüksek performansa sahiptir.
- Yeterli düzey: Konu ile ilgili yeterli bilgiye sahiptir.
- Temel düzey: Sadece temel bilgi ve becerilere sahiptir.
- Yetersiz düzey: Temel bilgi ve becerilere sahip değildir.

Performans düzeyi, öğrencinin bu seviyeye ulaşabilmesi için yapması gerekenlerin tanımlandığı noktadır. Performans düzey tanımlarının en doğru şekilde yapılması için, ilgili alanda uzman ve ilgili düzeydeki öğrencileri iyi tanıyan eğitimcilerin yardımına başvurulmalıdır (Çetin, 2011).

Performans düzeyi tanımlama süreci kesme puanı belirleme süreci başlamadan tamamlanmak zorundadır. Performans düzey tanımlarını yapan kişilerle, kesme puanı belirleyen kişilerin aynı uzmanlar olması zorunluluğu olmamakla birlikte, yeterli zaman ayırdıkları müddetçe, aynı uzmanlar hem performans düzeylerini tanımlayabilir hem de hem de kesme puanı belirleyebilir. Performans düzeylerini tanımlamada cinsiyet, etnik köken ve ırksal farklılıklar dikkate alınmalı, farklı türdeki okullar bu değerlendirmeden etkilenecekse bu okul türlerinin temsilcisi uzmanlardan yardım alınmalıdır (Zieky ve Perie, 2004).

Tablo 2-1’de ileri, yeterli, temel, temele yakın ve yetersiz olmak üzere beş farklı performans düzeyi ve bu performans düzeylerine ulaşmak için sınavı alan kişilerin 50 maddelik bir testten doğru yanıtlamaları gereken madde sayısını gösteren bir örnek verilmiştir.

Tablo 2-1. Farklı beş performans düzeyi için kesme puanı örneği (Çetin, 2011).

Performans düzeyi	Doğru yanıtlanan madde sayısı (N=50)	Kesme puanı
İleri düzey	40-50	40
Yeterli düzey	30-39	30
Temel düzey	20-29	20
Temele yakın düzey	10-19	10
Yetersiz düzey	0-9	-

Tablo 2-1 incelendiğinde, 5 farklı düzey belirleme çalışması yapılabilmesi için 4 farklı kesme puanının belirlenmesi gerektiği görülmektedir. Tabloya göre, sınavı alan kişi temele yakın düzey kategorisine yerleşmek istiyorsa bu düzeyin kesme puanı olan 10 adet maddeyi doğru yanıtlamalı, aynı şekilde yeterli düzey kategorisine yerleşmek istiyorsa, bu düzeyin kesme puanı olan 30 adet maddeyi doğru yanıtlamalıdır. Eğer sınavı alan kişiler başarılı ve başarısız şekilde iki kategoriye ayrılmak istenirse, 2 adet performans düzeyi belirlenmeli ve 1 kesme puanı belirlenmelidir. Standart belirleme kavramını ve sürecini daha iyi anlayabilmek için bilinmesi gereken iki kavram minimum yeterlik düzeyi ve kesme puanıdır.

2.1.1. Minimum yeterlik düzeyi

Minimum yeterlik, testte ölçülen nitelik ile ilgili bir bireyin minimum düzeydeki bilgi ve becerisidir. Minimum yeterlik düzeyi atanmak, mezun olmak veya bir işe girmek vb. için gerekli olan düzeydir. Standart belirleme sürecinde görev alacak puanlayıcıların minimum yeterlik düzeyini çok iyi kavramaları gerekmektedir (Mills, Melican ve Ahluwaila, 1991).

2.1.2. Kesme puanı

Bireyleri istenilen, incelenen özellik doğrultusunda olan ya da olmayan şeklinde sınıflandırmak için uygun olan noktadır. Yeterli ya da yetersiz gibi ayrımlar yapabilmek için belirlenen sınır puanıdır (Öner, 1997).

2.2. Standart Belirleme Süreci

Kişileri performans düzeylerine göre ayıran performans standartlarını, test puanları cetvelinde rakamsal hale getirme sürecine standart belirleme süreci denir (Hambleton, 2001). Bu süreç sonunda nihai bir kesme ya da yeterlik olarak adlandırılabilen bir puana ulaşılır. Hambleton (2001), standart belirleme sürecinde takip edilmesi gereken adımları şu şekilde sıralamıştır:

- Büyük ve örnekleyici bir uygulama grubunun seçilmesi,
- Standart belirleme yönteminin seçilmesi ve ilgili materyallerin hazırlanması,
- Performans düzeyleri tanımlarının hazırlanması,
- Standart belirleme yönteminde karar alınacak uzmanların eğitilmesi,
- Kullanılan yöntemle göre madde puanları ya da katılımcı kararlarının toplanması,
- Katılımcıların bir araya gelip tartışmalarına olanak sağlanması ve madde güçlük değerleri, madde ayıricılık değerleri gibi özetleyici bilgilerin elde edilmesi,
- Katılımcıların başka bir puanlama turu oluşturmalarını sağlayarak 5. ve 6. basamakta olduğu gibi bilgileri toplamak ve karar vermelerine imkan sağlamak,
- Katılımcıların bilgilerini yeniden gözden geçirmelerine bir kez daha imkan sağlayarak kesme puanına ulaşmak,
- Standart belirleme sürecinin değerlendirilmesinin yapılması,
- Standart belirleme sürecinin ve performans standartlarının geçerlik kanıtlarının bulunmasıdır.

Standart belirleme sürecinde dikkate alınması gereken bazı önemli noktaları Hambleton (1997), şu şekilde sıralamaktadır:

- Önemli standart belirleme çalışmalarından önce saha çalışması yapmak, muhtemel hataları en aza indirmek açısından çok önemlidir.
- Uzmanların ilgili alanı temsil edip, standartları belirleyecek yeterli sayıda olmasına dikkat edilmelidir.
- Standart belirleme süreci uzmanların geçmişinden, testin uzunluğundan, test maddelerinin formatından etkilenmektedir. Deneyimsiz uzmanların eğitimi ve uzun testler süreci uzatabilmektedir.
- Testin en azından bir kısmı, uzmanlara sınav koşullarına benzer koşullar altında verilerek, uzmanların maddelerin güçlüğü hakkında daha gerçekçi bilgiler edinmesi sağlanarak, öğrencilerden beklentilerini doğru tespit etmeleri sağlanabilir.
- Uzmanların standart belirleme sürecinde eğitilmeleri, süreci anlayıp daha kabul edilebilir standartlar belirlemelerine yardımcı olacaktır.
- Uzmanların ilk verdikleri kararlar ile ilgili tartışmalarına imkan verilmesi, varsa yanlış anlama ve dikkatsizlik gibi hatalarının farkına varıp, ikinci kararlarının daha isabetli olmasına yol açacaktır.
- Standart belirleme sürecinin her aşaması belgelendirilebilmelidir.

Kullanılacak standart belirleme yönteminin seçilmesinde kullanılacak esasları Hambleton (2001) ve Reckase (2006) şu şekilde sıralamaktadır:

- Test maddelerinin yapısı dikkate alınmalıdır. Çoktan seçmeli testlerde daha çok Angoff yönteminin, yapılandırılmış cevap içeren testlerde ise İşaretleme yönteminin kullanılması uygun olacaktır.
- Güvenirliği düşük testlerin standart belirlemede kullanılması uygun olmayacaktır.
- Standart belirleme süreci bazı yöntemlerde çok uzunken, bazı yöntemlerde oldukça kısadır.

- Bir yöntem daha önceden denenmişse, tekrar kullanılacağı zaman saha çalışmasına gerek kalmayacak, hem zaman hem de maddi açıdan kayıp yaşanmayacaktır.
- Yöntemlerin geçerliği ile ilgili algılar da yöntem seçiminde önemlidir. Söz gelimi bazı araştırmacılar Angoff yöntemi ile ilgili kaygılar taşıırken, bazı araştırmacılar Karşıt Gruplar yöntemini daha eleştirel bulmaktadır.

2.3. Standart Belirleme Yöntemleri

Son yıllarda alan yazında otuzdan fazla standart belirleme yöntemi betimlenmesine karşın, bu yöntemleri genel anlamda iki ana başlık altında toplayabiliriz (Crocker ve Algina, 1986).

- Test maddelerine dayanan yöntemler (test merkezli)
- Testi alanların performansına dayanan yöntemler (performans merkezli)

Jaeger (1989)'e göre bu iki sınıflama arasındaki ayrım uzman kararlarında nelerin göz önüne alındığı ile ilgilidir. Test merkezli yöntemlerde uzman kararları testin içeriğine göre verilir. Test maddeleri ya ayrı ayrı incelenir veya testin tamamı bütünsel anlamda ele alınırken, performans merkezli yöntemlerde ele alınan konu ile ilgili beklenen yeterlik düzeylerine göre öğrencilerin yeterlilikleri hakkında kararlar alınır.

2.3.1. Test maddelerine dayanan yöntemler (test merkezli yöntemler)

Test merkezli yöntemler, test maddeleri ile ilgili uzmanların vermiş oldukları kararlara dayanır. Test içerisinde yer alan maddeler hem tek tek hem de bir bütün olarak önemlidir. Uzmanlar değerlendirmelerini madde temelinde yaparlar. Her bir uzman her bir madde için minimum yeterlik düzeyindeki kişiyi düşünerek değerlendirme yapmaktadır. Bu tür yöntemlerde odak noktası testin kendisidir. Öğrencilere ilişkin yapılan değerlendirme ikinci planda kalmaktadır. Uzmanların maddelere ilişkin yaptıkları değerlendirmeler belirleyicidir. En çok bilinen test merkezli yöntemler Angoff, Nedelsky, Ebel, Jeager, Yes/No ve İşaretleme yöntemleridir.

Angoff yöntemi: 1971 yılında William H. Angoff tarafından ortaya konulan yöntem, her tür madde türünün bulunduğu testlerde uygulanabilen ve kullanımı en yaygın standart belirleme yöntemidir (Kane, 1994). Angoff yöntemi benzer diğer yöntemlere göre sürecin uzmanlara açıklanması, verilerin toplanması ve analizi bakımından sık kullanılan, kolay bir yöntemdir (Berk, 1986; Mills ve Melican, 1988). Kesme puanı her uzmanın her bir madde için verdiği tahmini değerlerin hesaplanması ile belirlenir. Uzmanlar her maddeyi geçme-kalma sınırındaki öğrencilerin cevaplayabilme olasılığını tahmin ederler. Angoff yöntemi ile kesme puanı belirlemeye yönelik 4 uzmanın 5 maddelik bir test hakkında verdikleri kararlardan oluşan bir örnek Tablo 2-2’de verilmiştir.

Tablo 2-2. Angoff yöntemi ile kesme puanı belirleme örneği

Uzmanlar	Madde1	Madde 2	Madde 3	Madde 4	Madde 5	MGP
Uzman 1	50	60	60	80	90	68
Uzman 2	70	80	40	60	80	66
Uzman 3	60	60	50	80	70	64
Uzman 4	40	80	65	70	70	65
Nihai teste ait kesme puanı						65.75

Tablo 2-2’de verilen örnekte görüldüğü gibi Angoff yöntemi ile kesme puanı belirlenirken, her uzmandan geçme-kalma sınırında olan öğrencilerden oluşan sınır gruptaki 100 öğrenciden testin her bir maddesini kaç öğrencinin doğru yanıtlayacağını tahmin etmesi istenir. Uzmanların maddelere verdikleri tahmini değerlerin aritmetik ortalaması uzmanlara ait minimum geçme puanlarını (MGP) verir. Tüm uzmanlara ait MGP’lerin ortalaması ise teste ait nihai kesme puanını verir.

Angoff yönteminin uygulanmasındaki önemli bir nokta, uzmanlara test maddelerinin doğru cevaplarının verilir verilmemesi durumudur. Uzmanlara doğru cevabı vermek, maddelerin daha kolay algılanmasına ve kesme puanlarının daha yüksek çıkmasına sebep olabileceken, karşı taraftan uzmanlara doğru cevapların verilmemesi uzmanların kendilerinin de sınava tabi tutuldukları algısına kapılmalarına ve doğru cevabı bulmaya çalışırken standart belirleme sürecini uzatabilmelerine neden olabilir. Bu problemin çözümü için, kesme puanı belirleme sürecinden önce uzmanların

inceleyebilmeleri amacıyla sadece test maddelerinin verilmesi, kesme puanı tespiti için verecekleri karar sürecinde ise maddelerin doğru cevaplarıyla birlikte verilmesi uygun olabilecektir (Livinston ve Zieky, 1982).

Bu yöntemde uzmanlardan 100 öğrenciden kaç öğrencinin ilgili maddeyi doğru cevaplayabileceğini tahmin etmeleri istendiği gibi, 1 öğrencinin ilgili test maddesini yüzde kaç ihtimalle yapabileceği de sorulabilir. 0 ile 100 arasında olan bu ihtimal, sorular kolaylaştıkça daha büyük olacaktır. Angoff yöntemiyle elde edilen minimum geçme puanlarından bazıları çok yüksek ya da çok düşük olabilmekte ve hesaplama sonucunu büyük oranda etkileyebilmektedir. Bu gibi durumlarda kesme puanının hesaplanmasında ortalama yerine ortancanın kullanılması da bir seçenektir. Karar alınan uzman sayısına göre en yüksek ve en düşük minimum geçme puanı atıldıktan sonra ortalamanın alınması da hataları minimize edebilmek adına kullanılabilecek bir diğer seçenektir.

Nedelsky yöntemi: Leo Nedelsky tarafından 1954 yılında ortaya konulan yöntem, sadece çoktan seçmeli testlerde kullanılırken, yöntemin işleyişinde uzmanlardan her olası yanlış seçeneği incelemeleri istenmektedir. Nedelsky yönteminin altında yatan varsayıma göre sınır gruptaki öğrenci test maddesini cevaplandırırken yanlış olduğunu düşündüğü seçenekleri eler ve kalan seçeneklerden rastgele birini seçer (Livinston ve Zieky, 1982).

Uzmanlardan, geçti-kaldı sınırı olarak tanımlanan sınır gruptaki bir öğrencinin test maddesinin seçeneklerinden kaç tanesinin yanlış olduğunu bilerek eleyebileceğini tahmin etmesi istenir. Bir'in elenmeden kalan seçenek sayısına bölümü sınır gruptaki öğrencinin maddeyi doğru cevaplayabilme olasılığını verecektir. Tüm maddeler için verilen tahminler toplanarak her bir uzman için minimum geçme puanları hesaplanır. Minimum geçme puanlarının ortalaması ise nihai kesme puanı olarak hesaplanır. Tablo 2-3'te Nedelsky yöntemi ile kesme puanı belirlemeye yönelik bir örnek verilmiştir.

Tablo 2-3. Nedelsky yöntemi ile kesme puanı belirleme örneği

Maddeler											
	M 1	M 2	M 3	M 4	M 5	M 6	M 7	M 8	M 9	M 10	
Elenmeden kalan seçenek sayısı	2	4	2	3	1	5	3	1	5	1	
Doğru cevaplama ihtimali	½=.50	¼=.25	½=.50	1/3=.33	1/1=1	1/5=.20	1/3=.33	1/1=1	1/5=.20	1/1=1	5.31

Tablo 2-3'te 10 maddeden oluşan bir test için bir uzmanın vermiş olduğu kararlar görülmektedir. 5 seçenekli olduğu düşünülen maddelerden uzman 1. maddeye ait seçeneklerden 3 tanesini sınır grup öğrencisinin eleyebileceğini ve 2 seçeneğin elenmeden kalacağını düşünmüştür. Böylece doğru yanıt 2 seçenekten 1 tanesi olacağından, maddenin doğru yanıtlanma olasılığı $\frac{1}{2}=.50$ olacaktır. Tüm olasılıklar bu şekilde bulunup toplanarak 1. uzmana ait minimum geçme puanı tabloda görüldüğü gibi 5.31 olarak hesaplanmıştır. Bu hesaplama tüm uzmanlar tarafından yapıлып bulunan MGP'lerin ortalaması alınarak teste ait nihai kesme puanı hesaplanmaktadır.

Sadece çoktan seçmeli testlerde kullanılabilmesi ve uzmanların maddeler için sadece belli olasılık değerleri verebilmeleri yöntemin sınırlılıkları olarak görülebilir. Çünkü bir uzmanın herhangi bir test maddesinin yanıtlanabilme olasılığı için verebileceği tahmini değerler: Elenmeden kalan seçenek sayısı 5 ise .20, 4 ise .25, 3 ise .33, 2 ise .50 ve 1 ise 1.00'dır (Cizek ve Bunch, 2007). Bir teste ilişkin birden fazla kesme puanı elde edilmesi gereken durumlarda Nedelsky yöntemi kullanılmamaktadır. Geçme-kalma, başarılı-başarısız ya da yeterli-yetersiz olarak bir ayırım yapılması gereken durumlarda kullanılmaktadır (Tülübaş, 2009).

Ebel yöntemi: R. L. Ebel tarafından 1972 yılında ortaya konulmuş iki aşamalı bir yöntemdir. Genellikle sağlık bilimleri alanında ve sınıf içi başarı testlerinde standart belirleme amacıyla kullanılmaktadır (Cizek ve Bunch, 2007).

İki aşamalı bir süreçten oluşan yöntemde ilk olarak uzmanlar testteki her maddeyi “zorluk” ve “uygunluk” olmak üzere sınıflamaya tabi tutar. Zorluk kategorisi orta, kolay ve zor olmak üzere üç, uygunluk kategorisi ise gerekli, önemli, kabul edilebilir ve tartışılabilir olmak üzere dört alt gruba ayrılır. Toplamda her madde 3x4

tipi bir tabloya yerleştirilir. İkinci olarak her uzmanın 3x4 tipi tabloya yerleştirilen maddeler için her hücrede bulunan madde veya maddeler için, tahminde bulunması beklenir. Uzman ilgili hücrede bulunan madde/maddeleri geçme-kalma sınırında bulunan 100 kişiden kaçının doğru yanıtlayacağına dair tahminde bulur. Her hücrede bulunan madde sayısı uzmanın verdiği yanıtlama olasılığı ile çarpılır ve tüm hücreler için bulunan bu değerler toplanarak ilgili uzmana ait minimum geçme puanı hesaplanır. Tüm uzmanlar için elde edilen minimum geçme puanlarının ortalaması ise testin nihai kesme puanını verir. Zorluk düzeyinin sınıflandırılmasında uzmanlara yardımcı olmak adına soruların güçlük düzeylerinin uzmanlara verilmesi faydalı olabilecektir. Ebel yöntemi ile kesme puanının hesaplanmasına yönelik 20 maddelik bir test için yapılan işlemler Tablo 2-4'te gösterilmiştir.

Tablo 2-4. Ebel yöntemi ile “zorluk” ve “uygunluk” boyutlarının tespit edilmesini gösteren bir örnek

		ZORLUK		
		Kolay 1, 7 %90	Orta 6, 11, 19, 20 %60	Zor 3, 12 %40
UYGUNLUK	Gerekli			
	Önemli	10, 15, 18 %80	5, 13, 14 %60	9 %30
	Kabul edilebilir	8 %70	2, 16, 17 %50	-
	Tartışılabilir	-	-	4 %35

Tablo 2-4'te görüldüğü gibi tüm maddeler uygunluk ve zorluk bakımından 12 hücreden birine yerleştirilir ve o hücrede bulunan tüm maddeler için genel bir cevaplama yüzdesi yazılır. Tüm hücrelerin tamamlanması zorunlu değildir. Tablo 2-5'te Tablo 2-4'ten yararlanarak kesme puanı hesaplama süreci gösterilmektedir.

Tablo 2-5. Ebel yöntemi ile kesme puanının hesaplanması

Kategoriler	Doğru cevaplama olasılığı	Soru sayısı	Hücrenin toplam puanı
Gerekli			
Kolay	90	2	.90x2=1.80
Orta	60	4	.60x4=2.40
Zor	40	2	.40x2=.80
Önemli			
Kolay	80	3	.80x3=2.40
Orta	60	3	.60x3=1.80
Zor	30	1	.30x1=.30
Kabul edilebilir			
Kolay	70	1	.70x1=.70
Orta	50	3	.50x3=1.50
Zor	-	-	-
Tartışılabilir			
Kolay	-	-	-
Orta	-	-	-
Zor	35	1	.35x1=.35
Uzmanla ait minimum geçme puanı			12.05

Tablo 2-5'te görüldüğü gibi 12 hücreden her biri için doğru cevaplama olasılığı ile hücredeki soru sayısı çarpılmış, tüm hücreler için bulunan değerler toplanarak bir uzmana ait minimum geçme puanı hesaplanmıştır. Tüm uzmanların minimum geçme puanlarının ortalaması ise testin nihai kesme puanını vermektedir.

Yes/No yöntemi: Impara ve Plake tarafından 1997 yılında geliştirilen yöntem bir Angoff çeşitlemesidir. Uzmanlardan istenen, testteki her bir maddeyi inceleyip minimum yeterlik düzeyindeki bir öğrencinin ilgili test maddesini doğru cevaplayıp cevaplayamayacağını tahmin etmektir. Uzman, öğrencinin maddeye doğru cevap vereceğini düşünüyorsa 1, yanlış cevap vereceğini düşünüyorsa 0 şeklinde öğrencinin maddeyi cevaplama olasılığını puanlar. Bu puanların toplamı ise ilgili uzman için bulunan minimum kesme puanıdır. Tüm uzmanlardan elde edilen minimum geçme puanlarının ortalaması teste ait nihai kesme puanını verir.

Yes/No yönteminde sınır gruptaki 1 öğrencinin maddeyi doğru cevaplayıp cevaplamayacağı sorulurken, Angoff yönteminde 100 kişiden kaçının maddeyi doğru yanıtlayabileceği sorulmaktadır. Uzmanlar büyük bir gruptan ziyade tek bir öğrenci üzerinden olasılığın hesaplanmasının daha kolay olduğunu ifade etmişlerdir. Yes/No yönteminde iki boyuta inen tahminlerin geçerliliğinin daha büyük olduğu gözlenmektedir. Bu anlamda Yes/No yönteminin basitliği bir avantaj olarak düşünülebilir.

Jeager yöntemi: Angoff yöntemi ile benzerlikler taşıyan, madde temelinde değerlendirme yapmayı gerektiren bir yöntemdir. Uzmanlar değerlendirme yaparken her bir maddeyi, “Değerlendirilen herkes bu maddeyi doğru yanıtlayabilir mi?” sorusunu dikkate alırlar (Cizek, 2001). Elde edilecek olasılık değerinin .50 ile 1.00 arasında olması yöntemin bir sınırlılığı olarak görülmektedir.

İşaretleme yöntemi: Mitzel, Lewis ve Green tarafından 1996 yılında geliştirilmiştir. Angoff yönteminin bazı eksikliklerini gidermek isteyen araştırmacılar, hem çoktan seçmeli hem de yapılandırılmış yanıtlar içeren maddelerin bir arada bulunduğu testlerde kullanabilmek, uzmanların iş yükünü azaltmak ve test içeriğini performans düzeyi tanımları ile birleştirmek amacıyla bu yöntemi geliştirmişlerdir (Mitzel, Lewis ve Green, 2001).

İşaretleme yöntemindeki esas, Madde Tepki Kuramı (MTK) kullanılarak maddelerin haritalanmasıdır. Maddeler ölçek üzerindeki yerlerine madde güçlükleri (p) dikkate alınarak sıralanırlar. Sıralama en basit maddeden en zor maddeye doğru yapılır (Çetin, 2011).

Bu yöntemde uzmanların görevi Angoff ya da diğer test merkezli yöntemlerdeki uzmanların görevinden daha farklıdır. Burada uzmanlar her bir madde için yargıda bulunmak yerine, kesme puanının maddeler arasındaki yerini saptarlar. Uzmanlar maddeleri yeterli ve yetersiz olarak kategorilere ayırıp bir kesme puanı belirlerken bir olasılık kararı vermek zorundadırlar (Huyhn, 2000).

Diğer yöntemlere göre daha yenidir ve az bilinmektedir. Maddeler tek tek baz alınmak yerine, testin bütünü esas alınır. Test maddeleri MTK kullanılarak zorluk dereceleri esas alınarak MTK yetenek ölçeğinde kolaydan zora sıralanır. Bu ölçek üzerinde kesme puanının nerde olması gerektiği uzmanlar tarafından belirlenir. İleri düzey, yeterli düzey ve yetersiz düzey olmak üzere üç standart belirlendiğinde uzmanlar öncelikle her standart düzeyindeki kişilerin neleri yapabileceklerini tanımlar, daha sonra bunu karşılayabilecek kesme puanını hesaplar. Böylece bir performans düzeyini diğerinden ayırt edebilecek şekilde maddeleri işaretlemiş olurlar. Seçimler yapılırken birçok uzmanın görüşüne başvurulmalıdır. Her uzman kendi belirlediği kesme puanını sunduktan sonra, diğer uzmanların belirledikleri ile uyumsuzluk olup olmadığına

bakılmaktadır. Bu şekilde uzmanların performans standardı tanımlamalarında ortak kaniya varmaları sağlanarak işaretleme süreci tekrarlanır ve her düzey için bir kesme puanı belirlenmiş olur (Tanrıverdi, 2006).

2.3.2. Testi alanların performansına dayanan yöntemler (performans merkezli yöntemler)

Testi alanların performansına dayanan yöntemler testi alan bireylerin yetenekleri göz önünde bulundurularak uygulanır (Irwin, 2007). Test merkezli yöntemlerde de öğrenci merkezli yöntemlerde olduğu gibi puanlayıcıya ihtiyaç duyulur. Ancak öğrenci merkezli yöntemi uygulayan uzmanlar testteki maddeler hakkında değil de testi alan bireyler hakkında bir yargıya varırlar. Bu sebeple puanlayıcıların testi alan bireylerin bilgi, beceri ve yetenek düzeyleri hakkında bilgi sahibi olmaları gereklidir (Taşdelen, 2009). Sıklıkla kullanılan iki yöntem Sınır Gruplar yöntemi ve Karşıt Gruplar yöntemidir.

Sınır gruplar yöntemi: Livingston ve Zeiky (1989) yeterli ya da yetersiz sınıflandırmasının net olarak yapılamamasına dayanarak sınır grup yöntemini geliştirmişlerdir. Bu yöntemde uzmanlardan yeterli ya da yetersiz olarak iki sınıflandırma yapmaları istenmektedir. Her iki sınıflandırmada arada kalan, bir başka ifade ile ne tam yeterli ne de tam yetersiz olduğu düşünülen kişilerin yer aldığı yeni bir grup oluşturulur. Oluşturulan bu yeni sınır grubun ortalaması ya da ortancası belirlenerek kesme puan elde edilir.

Sınır grup olan geçme-kalma sınırındaki öğrencilerin puanı esas alınır. Uzmanlardan bilgi ve yetenek bakımından sınır grupta olan öğrencileri tahmin etmeleri istenir. Sınırdaki öğrenciler belirlendikten sonra uygulanan testin sonunda elde edilen puanların ortancası kesme puanı olarak belirlenir. Kullanımının ve açıklamasının kolay ve basit olması en önemli avantajı iken, sınır grup öğrencilerinin testi alan grupta küçük bir yüzdeyi oluşturuyor olması nedeniyle bu öğrencileri ayırt etmede zorlanabilme ihtimalleri dezavantaj olarak görülmektedir (Çetin, 2011).

Ortalama yerine ortancanın kullanılmasının nedeni, ortancanın uç değerlerden az etkilenmesidir. Uç değer alan öğrenciler zaten sınır grupta değildirler. Eğer sınır grup öğrencilerinin puanları iki grupta toplanıyorsa yöntem iyi işliyor, puanlar kümelenmiyor

ve dağılım gösteriyorsa yöntem iyi işlemiyor, diğer bir ifadeyle sınır grup öğrencileri iyi belirlenememiş demektir (Livingston ve Zieky, 1982).

Zıt gruplar yöntemi: Bu yöntem standart belirleme sürecinde farklı yetenek grupları arasında beklenen performans düzeyi ile var olan performans düzeyi karşılaştırılmakta ve karşılaştırmaya göre kesme puanı belirlenmektedir (Tanrıverdi, 2006). Uzmanlar tarafından “yeterli” ve “yetersiz” olduğu düşünülen kişilerden oluşan iki zıt gruba test uygulanır. Her bir puan düzeyindeki öğrencinin “yeterli” grupta yer alma olasılığı hesaplanır (Cizek ve Bunch, 2007).

Zıt gruplar yöntemi aslında daha önceden bilinen farklı bir yöntemin genişletilmiş hali olarak Berk tarafından 1976 yılında geliştirilmiştir. Herhangi bir kesme puanı belirleme işlemi yapılmadan önce değerlendirilecek olan kişiler konuya ilişkin bir teste tabii tutulur. Kesme puanı belirleyecek olan uzmanlardan, değerlendirilen kişilerin puanlarına göre bir geçme-kalma ya da yeterli-yetersiz şeklinde bir kestirimde bulunmaları istenmektedir. Uzmanlar tarafından yapılan bu kestirim işleminden sonra yeterli ya da yetersiz olarak belirlenen iki gruba ait puan dağılımları oluşturulduktan sonra her iki dağılımın birbirlerini kestikleri nokta kesme puanı olarak belirlenmektedir. Belirlenecek olan kesme puanının her bir gruba düşecek olan kişi sayısından etkilenmesi söz konusu olabilir. Özellikle yeterli ya da yetersiz olarak yapılan sınıflandırmadan herhangi birinde diğerine göre ciddi bir kişi sayısı farklılığı söz konusu olduğunda elde edilen kesme puanı gerçeği yansıtmamaktadır. Bu gibi durumlarda kesme puanının belirlenebilmesi için farklı istatistiksel işlemler uygulanabilmektedir. Bunlardan biri her iki grubun ortancasını belirleyip, ortancalar arasında kalan orta noktayı belirlemektir. Her iki ortanca arasındaki bu orta nokta kesme puanı olarak kullanılabilmektedir. Zıt gruplar yöntemi kullanıldığında sıklıkla tercih edilen ikinci bir istatistik de lojistik regresyon analizidir. İki dağılım arasındaki kesişme noktasını belirlemek amacı ile kullanılmaktadır (Tülübaş, 2009).

Örneğin, 9. sınıf bir matematik başarı testi için, kesme puanı belirleme çalışmasında ilk olarak öğrenci grupları özel olarak seçilmelidir. Birinci gruptaki öğrenciler matematik seviyesi iyi, takviye kurslar almış öğrencilerden oluşurken, ikinci gruptakiler daha düşük seviyedeki öğrencilerden oluşmalıdır. İlk grup yeterli, ikincisi yetersiz olması beklenen gruptur. Her iki gruba da test uygulandıktan sonra

öğrencilerin aldıkları puanlarla yeterlikleri karşılaştırılabilir. Yeterli olması beklenen grup öğrencileri yüksek puanlar etrafında yığılma gösterirken, yetersiz olması beklenen grup öğrencileri düşük puanlar etrafında yığılma göstermelidir. Bu durumda uzmanlar hangi puanın yeterli ve yetersiz öğrencileri en iyi ayıran nokta olduğunu bulabilirler. Bu nokta testin kesme puanı olarak belirlenir (Tanrıverdi, 2006).

2.4. İlgili Literatür

Bu bölümde sırasıyla alanyazında yurt içinde ve yurt dışında standart belirleme yöntemleri ve yöntemlerin karşılaştırılması ile ilgili yapılan araştırmalara ait bilgilere ve bu bilgiler sonucunda ulaşılan bulgulara yer verilmiştir.

2.4.1. Yurt içinde yapılan standart belirleme yöntemleri ile ilgili çalışmalar

Yurt içinde yapılan çalışmalar genellikle test merkezli standart belirleme yöntemlerinin birbirleri ile karşılaştırılması şeklinde olmuştur. Yapılan çalışmalar kronolojik sıraya uygun şekilde verilmiştir.

Tanrıverdi (2006)'nin, standart belirleme yöntemlerinin geçme puanları üzerine etkisini incelediği çalışmasında, öğrencilerin başarılı veya başarısız olma durumlarının farklı standart belirleme yöntemlerine göre farklılaşıp farklılaşmadığı araştırılmıştır. Yedinci sınıf öğrencilerine uygulanan Matematik ve Türkçe testleri ile hem öğrenci başarı puanları elde edilmiş hem de aynı testler kullanılarak üç farklı yöntemle göre kesme puanları hesaplanmıştır. Araştırma sonucunda, Türkçe dersi için, farklı standart belirleme yöntemleri geçme puanı üstünde puan alan öğrenci yüzdelerine göre ikişerli olarak karşılaştırıldığında, Angoff ve Yes/No yöntemleri arasındaki yüzdeler arasında 0,05 düzeyinde, diğer yöntemler arasında 0,01 düzeyinde fark olduğu görülmüştür. Matematik dersi için yöntemlerin ikişerli karşılaştırmalarının tümünde 0,01 düzeyinde farklılaşma görülmüştür. Araştırma ile öğrenciler hakkında verilecek kararlar için kullanılabilecek standart belirleme yöntemlerinden hangisinin seçileceği öğrenci grubuna, derse, kullanılacak testin yapısına ve öğretmenlere bağlı olarak değişebileceği ortaya çıkmıştır.

Taşdelen (2009)'in Nedelsky ve Yes/No standart belirleme yöntemleri ile elde edilen kesme puanlarını Genellenebilirlik Kuramı ile karşılaştırdığı çalışmasında, araştırma verileri olarak 2008 yılı ilköğretim 6.sınıf SBS (Seviye Belirleme Sınavı)'de yer alan ve 16 maddeden oluşan fen ve teknoloji soruları kullanılmış, 415 öğrenciden veri toplanmış, 40 fen ve teknoloji öğretmeni tarafından kesme puanı belirlenmiştir. Nedelsky yöntemi ile yapılan puanlamada maddeler arası farklılığın ortaya çıkabildiği ancak puanlayıcılar arasında kesme puanı belirlemede uyum olmadığı görülmüştür. Angoff yönteminde ise yapılan puanlamada maddeler arası farklılığın ortaya çıkabildiği ve puanlayıcılar arasında tutarlılık olduğu görülmüştür. İki yöntem arasında kesme puanlarına göre geçen öğrenci yüzdeleri arasında 0,05 manidarlık düzeyinde farklılık olduğu ve Nedelsky yöntemine göre geçen öğrenci yüzdesinin Yes/No yöntemine göre geçen öğrenci yüzdesine oranla daha büyük olduğu sonucuna ulaşılmıştır. Ayrıca Yes/No ve Nedelsky yöntemleri ile standart belirlerken 10 puanlayıcının yeterli olacağı sonucuna varılmıştır.

Tülübaş (2009) çalışmasında Angoff ve Sınır Grup yöntemlerini norm oluşturma sürecinde kullanmış, 333 üniversite öğrencisinden veri toplanmıştır. Angoff ve Sınır Grup yöntemleri ile hesaplanan kesme puanları Beck Depresyon ve Beck Umutsuzluk Ölçekleri için oluşturulan normlarla karşılaştırılmıştır. Angoff yöntemi ile belirlenen kesme puanı Beck Umutsuzluk Ölçeği için normlardan farklı bulunmaz iken Sınır Grup yöntemi ile elde edilen kesme puanı Beck Depresyon Ölçeği için normlardan farklı bulunmuştur.

Çetin ve Gelbal (2010), standart belirleme yöntemlerinin geçme puanları üzerindeki etkisini inceledikleri çalışmalarında, 25'er maddelik Türkçe ve matematik testini kullanarak 125 7. sınıf öğrencisinden veri toplamış, 24 uzman kullanarak kesme puanı belirlemişlerdir. Öğrenciler hakkında verilecek kararlar için kullanılacak standart belirleme yöntemlerinden hangisinin seçileceğinin öğrenci grubuna, derse, kullanılacak testin yapısına ve öğretmenlere bağlı olarak değişebileceğini, Türkçe dersi için Nedelsky ve Yes/No yöntemleri arasında manidar bir fark olduğu, matematik dersi için Nedelsky ve Yes/No yöntemleri arasında manidar bir fark olmadığı sonucuna ulaşmışlardır.

Ömür ve Selvi (2010), araştırmalarında Angoff, Ebel ve Nedelsky yöntemleriyle farklı uzman sayıları kullanılarak elde edilen kesme puanlarının sınıflama tutarlığı indekslerini incelemiş, 17 maddeden oluşan çoktan seçmeli bir İngilizce başarı testini kullanarak 176 8. sınıf öğrencisi ve 30 uzmandan veri toplamışlardır. Araştırmacılar, uzman sayısına bağlı olarak kesme puanlarında önemli bir değişim gözlenmemekle birlikte üç yöntemde ve üç farklı uzman sayısında en yüksek sınıflama tutarlığı indeksinin Nedelsky yöntemi ile elde edildiği sonucuna varmışlardır.

Çetin (2011) çalışmasında, İşaretleme yöntemi ile Angoff yöntemini karşılaştırmış, 55 maddelik İngilizce dilbilgisi ve sözcük bilgisini ölçen bir başarı testi ile veri toplamıştır. Bu amaçla bir vakıf üniversitesinin hazırlık sınıfı öğrencilerine uygulanan dil sınavına yönelik İşaretleme ve Angoff yöntemlerine uygun şekilde 17 uzmandan kararlar alarak kesme puanları belirlemiştir. Bu sınav için mevcut yeterlik puanı ile belirlenen kesme puanlarını karşılaştırmıştır. Hem Angoff yöntemi ile hem de İşaretleme yöntemine ait bir parametrelili ve iki parametrelili tüm lojistik modellerde mevcut standarttan daha küçük bir kesme puanı elde edildiği görülmüştür. Üniversitenin belirlediği mevcut kesme puanı üzerinde puan alan öğrenci yüzdesi ise işaretleme yöntemine göre belirlenen iki kesme puanından RP50 (yeterlilik göstergesi olarak maddeleri cevaplama olasılığı %50 alındığında belirlenen kesme puanı, RP: Pespone Probability)'ye, RP67 (yeterlilik göstergesi olarak maddeleri cevaplama olasılığı %67 alındığında belirlenen kesme puanı)'ye ve Angoff yöntemine göre geçen öğrenci yüzdeleri arasında 0,01 düzeyinde fark olduğu belirlenmiştir. Cevaplama olasılıkları RP50 ve RP67'ye göre, İşaretleme yöntemi ile Angoff yöntemine göre belirlenen kesme puanları üzerinde puan alan öğrenci yüzdeleri arasında ise önemli bir fark olmadığı saptanmıştır.

Gündeğer (2012)'in çalışmasında Angoff, Yes/No ve Ebel standart belirleme yöntemleri ile elde edilen kesme puanları ve uzmanlardan elde edilen kararlar karşılaştırılmıştır. 20 maddeden oluşan bir matematik başarı testi ile 489 4. sınıf öğrencisinden ve 17 uzmandan veri toplanmıştır. Araştırma sonucunda üç farklı standart belirleme yöntemine göre başarılı sayılan öğrenci yüzdeleri arasında anlamlı fark olduğu görülmüştür. Yöntemlerden elde edilen kesme puanları ile öğrenciler başarılı-başarısız şeklinde kategorize edildiğinde Angoff ve Yes/No yöntemleri ile

Angoff ve Ebel yöntemleri arasında yüksek dereceli bir uyum, Yes/No ve Ebel yöntemleri arasında orta dereceli bir uyum olduğu görülmüştür.

Taşdemir (2013) çalışmasında, sınıf öğretmenliği öğrencilerinin test başarılarının Yes/No ve Nedelsky standart belirleme yöntemleri ve sınır değer saptama yöntemleri İşlem Karakteristik Eğrisi (IKE) ve Aralık Tahmini ile belirlenen sınıflama doğruluklarını incelemiştir. Veriler 30 maddelik 2006-2012 yılları arasında KPSS (Kamu Personeli Seçme Sınavı)'de sorulmuş Türkçe ve matematik sorularını içeren bir başarı testi ile toplanmış; 211 öğrencinin puanları ve 40 uzmanın kararları araştırmada kullanılmıştır. Bu araştırmanın sonucunda, en yüksek kesme puanının Türkçe ve matematik testi için Yes/No yöntemi ile elde edildiği, başarılı sayılan öğrenci yüzdeleri arasında manidar farklılık olduğu görülmüştür. Yes/No yönteminde Türkçe ve matematik maddelerine ait uzman kararları arasında uyum olmadığı, Nedelsky yöntemine göre kesme puan belirlemede uzman kararları arasında uyumun olduğu görülmüştür.

2.4.2. Yurt dışında yapılan standart belirleme yöntemleri ile ilgili çalışmalar

Yurt dışında yapılan standart belirleme yöntemleri ve yöntemlerin karşılaştırılmasına yönelik çalışmaların yurt içinde yapılan çalışmalara oranla çok daha fazla olduğu görülmektedir. Araştırmaların genel olarak test merkezli yöntemlerden Angoff, Yes/No ve Nedelsky, öğrenci merkezli yöntemlerden Karşıt Gruplar ve Sınır Grup yöntemleri üzerine yoğunlaştığı görülmektedir.

Andrew ve Hecht (1976) çalışmalarında, Ebel ve Nedelsky yöntemlerini iki farklı uzman grubu kullanarak karşılaştırmıştır. İki uzman grubundan her iki yöntem için de kararlar alınmış, grupların belirlediği kesme puanlarının hem Ebel yönteminde hem de Nedelsky yönteminde farklılık göstermediği görülmüştür. Ancak aynı uzman grubunun Ebel yöntemi ve Nedelsky yöntemiyle elde edilen kesme puanları karşılaştırıldığında yöntemlere göre puanlar arasında anlamlı bir farklılığın olduğu görülmüştür.

Brennan ve Lockwood (1980) çalışmalarında, Angoff ve Nedelsky standart belirleme yöntemlerine ait kesme puanlarını Genellenebilirlik kuramını kullanarak karşılaştırmışlar ve yöntemlerin belirlediği kesme puanları arasında anlamlı bir

farklılığın olduğu sonucuna ulaşmışlardır. Bu farkın nedeni olarak, bu iki yönteme ait sürecin farklı olmasını ve puanlayıcıların sınır gruba ait minimum yeterlik düzeyini algılamadaki farklılıklarını görmüşlerdir.

Skakun ve Kling (1980) çalışmalarında Nedelsky yöntemi ile Ebel yönteminin iki varyasyonu Ebel 1 ve Ebel 2 yöntemlerini karşılaştırmışlardır. Maddelerin güçlük düzeyleri bakımından sınıflandırılmasını temel alan Ebel 1 ve maddelerin önem düzeyleri bakımından sınıflandırılmasını temel alan Ebel 2 yöntemlerine göre belirlenen kesme puanlarının büyük farklılıklar göstermediği, Nedelsky yönteminin ise bu iki yöntemden oldukça küçük kesme puanı belirlediği sonucuna ulaşılmıştır.

Harasym (1981) analizlerinde Rash modelinden faydalandığı çalışmasında, Yes/No yöntemi ile Nedelsky yöntemini karşılaştırmak amacıyla 3 paralel testi 2. sınıfta öğrenim gören 3 sınıfa uygulamış, araştırma sonucunda iki yönteme ait süreçlerin, standart belirlemede ve öğrenci başarısı üzerinde anlamlı etkisi olduğu sonucunu bulmuştur.

Poggio ve diğerleri (1982) dört farklı yöntemi karşılaştırdıkları araştırmalarının sonucunda, Angoff, Ebel, Nedelsky ve Karşıt Gruplar standart belirleme yöntemleri ile elde edilen kesme puanları arasında büyük farklılıklara rastlanmadığını, en yüksek kesme puanının Ebel yöntemi tarafından belirlendiği, onu da sırasıyla Angoff, Karşıt Gruplar ve Nedelsky yöntemlerinin izlediği bulmuşlardır.

Livingston ve Zieky (1983) çalışmalarında Karşıt Gruplar yöntemi, Sınır Grup yöntemi, Angoff ve Nedelsky yöntemlerini okuma ve matematik testleri aracılığıyla karşılaştırmış, Karşıt gruplar ve Sınır Grup yöntemlerinin birbirine çok yakın kesme puanı belirlerken, dört yöntem arasında büyük farklılıkların olmadığı sonucuna ulaşmışlardır.

Mills (1983)'in Angoff, Karşıt Gruplar ve Sınır Grup yöntemlerini karşılaştırdığı çalışmasında, Angoff ve Karşıt Gruplar yöntemlerinin benzer sonuçlar verirken, Sınır Grup yönteminin diğer yöntemlerden oldukça farklı kesme puanı belirlediği görülmüştür.

Mills ve Berr (1983)'in Ebel, Angoff ve Karşıt Gruplar yöntemini karşılaştırdıkları araştırmanın sonucunda Ebel ve Angoff yöntemleri ile elde edilen kesme puanlarının, Karşıt Gruplar yöntemi ile elde edilen kesme puanından daha yüksek olduğu görülmüştür.

Appleton (1984)'in Nedelsky, Ebel ve Karşıt Gruplar standart belirleme yöntemlerini karşılaştırdığı çalışmasının sonucunda, Nedelsky yöntemi ile Karşıt Gruplar yöntemine ait kesme puanlarının birbirine yakın olduğu, Ebel yönteminin daha yüksek bir kesme puanı belirlediği görülmüştür.

Norcini ve diğerleri (1987) tarafından yapılan çalışmada, farklı zamanlarda verilen kararlar arasındaki ilişkileri ortaya çıkarmak amacıyla, karar verme süreçleri birer aylık aralıklarla buluşma öncesi (1. durum), buluşma (2. durum) ve buluşma sonrası (3. durum) olmak üzere üç farklı şekilde belirlenerek uzmanlardan bu üç zaman dilimi için Angoff yöntemi ile görüş alınmıştır. Birinci durumda verilen kararlara ait kesme puanının diğerlerinden daha düşük olduğu; ikinci durumda verilen kararlara ait kesme puanının ise diğerlerinden daha yüksek olduğu sonucuna ulaşılmıştır. Ancak üç sürece göre geçen öğrenci yüzdeleri arasında manidar fark bulunamamıştır.

Livingston ve Zieky (1989) yaptıkları çalışmada, dört okulda Sınır Grup, Karşıt Gruplar ve Nedelsky yöntemlerini, farklı dört okulda ise Sınır Gruplar, Karşıt Gruplar ve Angoff yöntemlerini karşılaştırmışlardır. Araştırma sonucunda, Nedelsky ve Angoff yöntemlerinin tutarlı sonuçlar vermediği görülmüştür. Sınır Gruplar ve Karşıt Gruplar yöntemleri ile yeterli ve yetersiz olarak kategorize edilen öğrenci gruplarındaki öğrenciler eşit sayıda olduğunda birbirine benzer sonuçlar verirken, yeterli öğrenci sayısının daha fazla olduğu durumlarda Karşıt Gruplar yönteminin kesme puanının Sınır Grup yöntemine göre daha düşük olduğu görülmüştür. Yetersiz öğrenci sayısı fazla olduğunda ise Karşıt Gruplar yöntemi ile elde edilen kesme puanı Sınır Grup yönteminin kesme puanına göre daha yüksek çıkmıştır.

Impara ve Plake (1997)'in Angoff ve Yes/No yöntemlerini karşılaştırdıkları çalışmalarında karar verme süreci iki farklı durumda gerçekleşmiştir. Birinci durumda uzmanlar kendi kararlarını verirken; ikinci durumda uzmanlara birinci durumda verilen kararlara göre kalan öğrenci yüzdeleri, madde güçlükleri gibi istatistikler verilerek karar

vermeleri sağlanmıştır. İki durum ile elde edilen kesme puanları arasında Angoff yönteminde 8 puanlık, Yes/No yönteminde de yaklaşık 1 puanlık bir düşüş olduğu görülmüştür. Ayrıca ikinci durumda standart sapma ve varyansın azaldığı, bir başka ifade ile uzmanların kararları arasındaki değişkenliğin azaldığı görülmüştür.

Impara ve Plake (1998) yaptıkları araştırmada 26 ilköğretim 6.sınıf fen bilgisi öğretmeninin, Angoff yöntemine göre 50 çoktan seçmeli test maddesine verdikleri, oran tahminlerindeki başarılarını incelemişlerdir. Öğretmenlerin, birisi 6. sınıf için sınır grup öğrencilerinin performansı, diğeri de sınıfın tamamının performansı olmak üzere, iki öğrenci grubunu dikkate alarak verdikleri kararların incelenmesi sonucunda, öğretmenlerin sınıfın tamamı için yaptıkları tahminlerin, sınır gruba göre daha doğru olduğu ancak her iki grup için de yeterince yüksek olmadığı görülmüştür.

Chang (1999) çalışmasında Nedelsky ve Angoff standart belirleme yöntemlerini karşılaştırmış, bu amaçla Araştırma Metotları dersi için kesme puanı belirlemede 80 eğitim fakültesi mezunu öğrenciden görüş almıştır. Araştırma sonucunda Nedelsky yönteminin, Angoff yöntemine göre daha düşük kesme puanları belirlediği ve Nedelsky yönteminde uzmanlar arası uyumun daha yüksek düzeyde olduğu görülmüştür. Nedelsky yöntemi ile belirlenen kesme puanlarının, öğrencilerin gerçek puanlarına daha yakın olduğu, Angoff yöntemi ile belirlenen kesme puanının öğrencilerin gerçek performanslarından büyük oranda saptığı gözlenmiştir.

Hurtz ve Hertz (1999), çalışmalarında Angoff yönteminde kaç uzman kullanılması gerektiğini genellenebilirlik teorisini ve sekiz farklı mesleki lisans sınavı verilerini kullanarak araştırmışlar ve yaklaşık 10-15 uzmanın standart belirlemede yeterli olacağı sonucuna ulaşmışlardır.

Impara, Plake ve Irwin (2000), Angoff yönteminde uzmanlar tarafından yapılan madde performans tahmininin uzmanlar içi ve uzmanlar arası güvenilirliğini analiz etmek amacıyla gerçekleştirdikleri çalışmalarının sonucunda, madde performans tahminlerinin uzmanlar içi ve uzmanlar arası tutarlı olduğuna ilişkin kanıtlar bulmuşlardır.

Buckhendal, Plake, Impara ve Smith (2002) tarafından yapılan araştırmada Angoff yöntemi ile İşaretleme yöntemi 7. sınıf matematik testi kullanılarak

karşılaştırılmış, araştırma sonucunda matematik dersine ilişkin kesme puanları her iki yöntemle göre de belirlenmiş ve elde edilen kesme puanları arasında büyük farklılıklar gözlenmemiştir.

Wang, Pan ve Austin (2003)'in, çeşitli standart belirleme yöntemleri ve bu yöntemlerin geçme puanları üzerindeki etkilerini inceledikleri çalışmalarının sonucunda, kullanılan on iki standart belirleme yöntemine göre belirlenen geçme yüzdelерinin yöntemler arasında, hatta aynı kategoride yer alan yöntemlerin bile kendi içlerinde oldukça fazla değiştiğini gözlemlenmiş, bu değişimin %29'dan %79'a kadar değiştiği sonucuna ulaşılmıştır.

Chang, Van der Linden ve Vos (2004), yeni bir test merkezli standart belirleme yöntemi geliştirmek ve yöntemin uzmanlar arası tutarsızlıklarını tespit etmek amacıyla yaptıkları araştırmada, uzmanlar alternatif cevaplarla standart belirleme yöntemini değerlendirmek için çoktan seçmeli testteki maddelerin çözümünde kullanılan süreci değerlendirmişlerdir. Nedelsky ve Angoff yöntemleri ile karşılaştırıldığında, yeni yöntemle belirlenen kesme puanlarının Nedelsky yöntemi ile belirlenen kesme puanına göre daha yüksek, Angoff yöntemiyle belirlenen kesme puanına göre daha düşük olduğu görülmüştür.

Downing ve diğerleri (2006) yaptıkları araştırmada Sınır Grup, Karşılaştırmalı Gruplar, Angoff, Hofstee ve Ebel yöntemlerinin süreçleri üzerinde çalışmışlar, farklı standart belirleme yöntemlerinin farklı standartlar ürettiği, en iyi standardın olmadığı, uzman kararlarının öznel olduğu ve tüm standartların politik kararlar olduğu sonucuna varmışlardır.

Irwin (2007) çalışmasında Karşıt Gruplar 1 ve Karşıt Gruplar 2 adını verdiği yöntemlerle; Angoff, Karşıt Gruplar ve Sınır Gruplar yöntemlerini karşılaştırmış ve Karşıt Gruplar 1 ve Karşıt Gruplar 2 yöntemleri ile elde edilen kesme puanlarının Angoff ve Karşıt Gruplar yöntemlerinden oldukça yüksek; Sınır Grup yönteminden ise düşük olduğu sonucuna ulaşmıştır. Karşıt Gruplar 1 ve Karşıt Gruplar 2 yöntemlerinin belirlediği kesme puanlarının ise aynı veriler üzerinde sürekli değiştiği görülmüştür.

Hess, Subhiyah ve Giordano (2007), kümeleme analizi ile Nedelsky, Ebel ve Yes/No yöntemlerini bir arada kullanmış ve bireyler hakkında verilen geçti-kaldı

kararlarının doğruluklarını incelemiştir. 652 kişiden toplanan yanıtlara göre sınava girenler gruplanmış, Ebel ve Nedelsky yöntemleri için uzman kararlarına göre belirlenen geçti-kaldı sınıflamaları benzer bulunurken, Angoff ve kümeleme analizi yöntemleri ile belirlenen kesme puanlarının yüksek derecede birbiri ile tutarlı olduğu görülmüştür.

Clauser ve arkadaşları (2009), çalışmalarında Angoff yöntemi için uzmanlarca belirlenen kesme puanlarının kararlılığını ölçmek amacıyla genellenebilirlik teorisinden yararlanmışlardır. Araştırmanın sonucunda, tartışmaların uzman kararları arasındaki farklılığı azalttığı, maddeler için yapılan uzman tahminlerinde ise çeşitliliğin arttığı gözlenirken, uzman tahminleri ile maddelerin gerçek güçlükleri arasındaki ilişkiyi güçlendiremediği görülmüştür. Ayrıca uzmanların zor maddeler için başarı oranını var olandan daha fazla tahmin ederlerken, basit maddeler için cevaplama olasılığını var olandan daha düşük oranlarda tahmin ettikleri görülmüştür.

BÖLÜM III

3. Yöntem

Bu bölümde, araştırmanın türü, verilerin toplandığı çalışma grubu, veri toplama araçları ve veri analizlerinde kullanılan istatistiksel yöntemlere yer verilmiştir.

3.1. Araştırmanın Modeli

İki ya da daha çok sayıdaki değişken arasındaki birlikte değişimin varlığını ve/veya derecesini belirlemeyi amaçlayan araştırma modelleri, ilişkisel türden araştırma modelleri olarak tanımlanır (Karasar, 1982, s. 85). Bu araştırma ile farklı üç standart belirleme yöntemine göre elde edilen kesme puanları karşılaştırılmıştır. Bu açıdan çalışma, ilişkisel türden bir araştırmadır. Olayların, objelerin, varlıkların, kurumların, grupların ve çeşitli alanların “ne” olduğunu açıklamaya çalışan araştırmalar betimsel araştırmalardır (Kaptan, 1977, s. 129). Araştırma ile Angoff, Nedelsky ve Ebel yöntemlerinin belirlediği kesme puanları incelendiğinden ve bu durum saptamaya yönelik olduğundan, çalışma bu yönüyle betimsel bir araştırma niteliği taşımaktadır.

3.2. Verilerin Elde Edildiği Grup

Araştırmada, verilerin elde edildiği iki farklı grup bulunmaktadır. Birinci grup öğrencilerden, 2. grup ise uzmanlardan oluşmaktadır. Öğrenci grubu, Bolu ilinin 13 farklı lisesinin 9., 10., 11. ve 12. sınıflarında öğrenim görmekte olan toplam 245 öğrenciden oluşmaktadır. Öğrenci seçimi tamamen gönüllülük esasına dayalı olarak amaçsal örnekleme yöntemine göre yapılmıştır. Öğrencilere araştırmacı tarafından geliştirilen 9. sınıf 1. dönem matematik müfredatında bulunan kümeler ve denklem-eşitsizlikler öğrenme alanlarını içeren 26 maddelik bir matematik başarı testi

uygulanmıştır. Aşağıda verilen Tablo 3-1’de araştırmaya katılan öğrencilerin okulları ve öğrenci sayıları verilmektedir.

Tablo 3-1. Araştırmaya katılan okullar ve okulların öğrenci sayıları

Araştırmaya Katılan Okullar	Katılan Öğrenci Sayısı
Merkez Teknik ve Endüstri Meslek Lisesi	47
Fen Lisesi	56
Emine Mehmet Baysal Anadolu Lisesi	61
Güzel Sanatlar Lisesi	12
Bolu Anadolu Lisesi	7
İmam Hatip Lisesi	5
Mimar Teknik ve Endüstri Meslek Lisesi	25
Anadolu Öğretmen Lisesi	6
İzzet Baysal Anadolu Lisesi	7
Turizm Otelcilik Meslek Lisesi	4
Mustafa Çizmeci Anadolu Lisesi	4
Bolu Atatürk Anadolu Lisesi	6
Canip Baysal Anadolu Lisesi	5
TOPLAM	245

Karar verme aşamasında görüşlerine başvurulacak uzman grubun sayısı standart belirleme yöntemlerinde en çok tartışılan konulardan biridir. Uzman sayısına ilişkin farklı görüşler vardır. Jeager (1989)’a göre uzman sayısı arttıkça, standart hata azalacaktır. Fakat Jeager (1989) kullanılan standart belirleme yöntemlerinde uzmanların uzlaşması gerektiği durumlarda, uzman sayısının fazla olmasını bir dezavantaj olarak görmektedir. Her ne kadar uzman sayısına bir sınırlama getirilmese de 10 ile 15 arasında olan bir uzman sayısının yeterli olduğu ileri sürülmektedir.

Verilerin elde edildiği 2. grup olan uzmanlar grubu toplam 19 kişiden oluşmaktadır. Başlangıçta her bir yöntem için kesme puanı belirlemek amacıyla 21 uzmandan karar alınmıştır. Ancak bir uzmanın, Ebel yöntemiyle kesme puanı belirleme sürecinde matematik başarı testinin bazı maddelerini “Uzman Görüşleri Formu”na eksik yerleştirdiği, bir uzmanın ise Nedelsky yöntemiyle kesme puanı belirleme sürecinde

elenebilecek seçenek sayısını, elenemeyecek seçenek sayısı olarak düşünüp yanlış karar verdiği görülmüştür. Bu nedenle iki uzman değerlendirme dışı tutularak toplam 19 uzmandan kesme puanı belirleme sürecinde karar alınmıştır. Her bir uzman için Angoff, Nedelsky ve Ebel yöntemlerinin her biri için bir minimum geçme puanı hesaplanmıştır. Uzman seçimi amaçsal örnekleme yöntemiyle, gönüllülük esasına göre yapılmıştır. Uzman grubu alanında deneyimli matematik öğretmenlerinden oluşmaktadır.

3.3. Verilerin Toplanması ve Veri Toplama Araçları

Veriler öğrenciler ve uzmanlar olmak üzere iki farklı gruptan toplanmıştır. Her iki gruptan da toplanan veriler için araştırmacı tarafından geliştirilen bir başarı testi kullanılmıştır. Öğrencilerden toplanan veriler bu başarı testinden (Ek 1) alınan öğrenci puanlarıdır. Öğrencilerden veri toplama aşamasında, 9. Sınıflar Matematik Dersi Öğretim Programı 1. dönem müfredatında bulunan Kümeler ile Denklem ve Eşitsizlikler öğrenme alanlarına ait 17 kazanımın her biri için 3 adet olmak üzere toplam 51 tane madde başarı testi geliştirmek amacıyla araştırmacı tarafından hazırlanmıştır. Hedefle içerik arasındaki ilişkinin belirlenmesi amacıyla araştırmacı tarafından hazırlanan belirtke tablosu ile beraber hazırlanan başarı testi, kapsam geçerliğinin sağlanabilmesi ve maddelerde olabilecek maddi hataların tespit edilebilmesi açısından, alanında uzman 3 matematik öğretmenine inceletilmiştir. Aynı anda iki uzmanın ilgili kazanımı ölçmediği gerekçesi ile uygun görmediği 12. ve 24. maddeler testten çıkarılmıştır. Uzmanların çeşitli nedenlerle bireysel olarak uygun görmediklerini düşündükleri 3., 6., 14., 39., 46., maddeler de testten çıkarılmıştır. Araştırmacı tarafından, testin bir ders saatinde yapılabilirliğini sağlamak ve her bir kazanım için iki maddenin yazılması düşüncesi ile 8., 16., 18., 22., 27., 29., 34., 36., 42. ve 49. maddeleri de atılarak 17 kazanımı ölçen 34 maddelik bir matematik başarı testi hazırlanmıştır.

Hazırlanan başarı testi Bolu ilinde 13 farklı lisede bulunan 490 9., 10., 11. ve 12. sınıf öğrencisine bir ders saati süre verilerek uygulanmıştır. Uygulama gönüllülük esasına göre yapılmış, yanıtlamak istemeyen öğrencilere test verilmemiştir. Tüm öğrencilerin maddelerin tamamını okuyabilmeleri adına, bir ders saati sonunda her maddeye ulaşamayan öğrencilere bir ders saati ek süre verilmiştir.

Denemelik test 490 öğrenciye uygulanmış, uygulama sonucu elde edilen veriler Excel'e aktarılmıştır. 490 öğrenciden oluşan çalışma grubu tekler ve çiftler olarak ikiye ayrılmıştır. Teklerden oluşan ilk yarıdan elde edilen veriler ITEMAN (Version 3.5) programı ile madde analizine tabi tutulmuş ve madde seçimi yapılmıştır. Madde seçiminde ilk olarak madde ayırt edicilik indeksi (r_{jx}), sonra madde güçlük indeksi (P_j) göz önünde bulundurulmuştur. Aşağıda bulunan Tablo 3-2'de çalışma grubunun ilk yarısından elde edilen, denemelik testin madde istatistikleri verilmiştir.

Tablo 3-2. Deneme uygulaması sonucu oluşan madde istatistikleri

Madde	Güçlük indeksi (P_j)	Ayırt edicilik indeksi (r_{jx})	Madde	Güçlük indeksi (P_j)	Ayırt edicilik indeksi (r_{jx})
1	0,71	0,43	18	0,53	0,82
2	0,78	0,46	19	0,27	0,79
3	0,73	0,52	20	0,44	0,63
4	0,70	0,51	21	0,42	0,78
5	0,64	0,67	22	0,49	0,84
6	0,44	0,72	23	0,41	0,86
7	0,57	0,72	24	0,43	0,79
8	0,41	0,85	25	0,55	0,77
9	0,51	0,81	26	0,72	0,53
10	0,30	0,92	27	0,62	0,66
11	0,54	0,73	28	0,68	0,62
12	0,42	0,71	29	0,71	0,65
13	0,42	0,64	30	0,50	0,85
14	0,45	0,76	31	0,54	0,80
15	0,24	0,71	32	0,79	0,25
16	0,42	0,67	33	0,64	0,74
17	0,47	0,82	34	0,50	0,75

r_{jx} : Çift serili korelasyon katsayısı

Tablo 3-2'de görüldüğü gibi maddelerin ayırt edicilik indeksleri 0,25 ile 0,92 arasında, madde güçlük indeksleri ise 0,24 ile 0,79 arasında değişirken, KR-20 iç

tutarlık güvenirlik katsayısı 0,94 gibi yüksek bir değer bulunmuştur. Osterlind (1992), başarı testleri için testteki maddelerin güçlük indekslerinin dizi genişliğinin 0,40 ile 0,80 arasında, Özçelik (1992) ise 0,20 ile 0,80 arasında olması gerektiğini belirtmektedir. Atılgan (2011)'a göre madde ayırt edicilik indeksi 0,40 ve daha yüksek olan maddeler çok iyi işleyen maddelerdir ve teste olduğu gibi alınabilir. Öğrenci başarısını belirlemek amacı ile geliştirilen bir başarı testinde hem kolay hem de zor sorular olmalı ve çoğunluğu orta güçlükte olmalıdır (Tekin, 1979; Turgut, 1993; Baykul, 2000). Bu bilgiler ışığında, 32. madde ayırt edicilik indeksinin 0,40 dan düşük olması, 10., 15. ve 19. maddeler de güçlük indekslerinin 0,40 dan düşük olmaları nedenleri ile testten çıkarılmıştır. İkinci, 4., 26. ve 29. maddeler de nispeten ayırt edicilik indekslerinin düşük, madde güçlük indekslerinin de düşük olması nedenleri ile, her kazanımdan en az bir soru olması koşulu da dikkate alınarak testten çıkarılmış ve 26 maddeden oluşan nihai test oluşturulmuştur. Denemelik testten seçilen maddelerden oluşturulan nihai testin kestirilen istatistikleri Tablo 3-3'te gösterilmiştir.

Tablo 3-3. Nihai testin kestirilen istatistikleri

	n	K	s	$\bar{X}$	$\bar{P}$	KR-20
Nihai Test	245	26	7,91	13,33	0,51	0,95

Tablo 3-3'te görüldüğü gibi denemelik testin uygulanması sonucunda testin iç tutarlık güvenirlik katsayısı 0,95 gibi yüksek bir değer kestirilmiştir. Büyüköztürk (2009)'e göre güvenirlik katsayısının 0,70 ve daha yüksek olması test puanlarının güvenirliği için genel olarak yeterli görülmektedir. Murphy ve Davidshofer (1991)'e göre ise sınıf içi başarı testlerinde iç tutarlık katsayısı en az 0,75 olmalıdır. Bu nedenle kestirilen güvenirlik değerinin (KR-20=0,95) yeterli olduğu söylenebilir. Testin 0,51 olarak bulunan ortalama güçlük değeri testin orta güçlükte bir test olduğunu göstermektedir.

Testin geçerliğini hesaplamak için öğrencilerin 9. sınıf 1. dönem sonu matematik karne notları ile testten aldıkları puanlar arasındaki ilişkiye Spearman Brown Sıra Farkları Korelasyon Katsayısı ile bakılmıştır. Matematik başarı testinin ölçüt geçerliği olarak da tanımlanabilecek 0,44 korelasyon değeri, orta derecede bir ilişkinin varlığını ortaya koymaktadır.

Geçerlik ve güvenirlik çalışmaları tamamlanan ve 26 maddeden oluşan nihai test 490 öğrencinin 2. yarısına karşılık gelen ve çiftler olarak adlandırılan gerçek gruba uygulanmış, nihai testin nihai uygulaması sonucu oluşan test istatistikleri aşağıda Tablo 3-4'te verilmiştir.

Tablo 3-4. Nihai testin uygulanması sonucu oluşan test istatistikleri

	n	K	s	$\bar{X}$	$\bar{P}$	KR-20
Nihai Test	245	26	7,73	13,32	0,51	0,93

Tablo 3-4 incelendiğinde testin iç güvenirlik katsayısının 0,93 gibi yüksek bir değer aldığı görülmektedir. Nihai testin nihai uygulama sonuçları, öğrencilerin 9. sınıf ara dönem matematik karne notları ile karşılaştırılmış, aradaki ilişkiye Spearman Brown Sıra Farkları Korelasyon Katsayısı ile bakılmıştır. 0,42 korelasyon değeri iki değişken arasında orta derecede bir ilişkinin varlığını göstermektedir. 0,51 bulunan testin ortalama güçlük değeri, nihai testin orta güçlükte olduğunu göstermektedir.

Uzmanlardan veri toplama aşamasında, 26 maddeden oluşan nihai testin her bir maddesini üç farklı standart belirleme yöntemine göre değerlendirmek ve her bir yöntem için kesme puanı hesaplamak amacıyla uzman görüşleri değerlendirme formu (Ek 2) hazırlanmıştır. Araştırmacı tarafından hazırlanan form yardımıyla alanında uzman 21 matematik öğretmeninden görüş alınmış ve her uzmanın her bir maddeye verdikleri kararlar doğrultusunda her bir yöntem için bir kesme puanı belirlemiştir. Ancak bir uzmanın Ebel yönteminin, bir uzmanın da Nedelsky yönteminin uygulanma sürecinde yaptığı maddi hatalar nedeniyle iki uzmanın kararları değerlendirme dışı tutulmuş, toplamda 19 uzmanın kararları kesme puanı belirleme sürecinde kullanılmıştır.

Uzmanlardan veri toplamadan önce araştırmacı öğretmenlere Angoff, Nedelsky ve Ebel standart belirleme yöntemleri hakkında bilgi vermiştir. Öğretmenlere minimum yeterlik düzeyi, sınır grup, kesme puanı gibi standart belirleme sürecinde kullanılan temel kavramlar hakkında açıklama yapılmıştır. Uzmanlardan testin her bir maddesini Angoff, Nedelsky ve Ebel standart belirleme yöntemlerine göre değerlendirmesi istenmiştir.

Angoff yöntemi ile yapılan 1. değerlendirmede 1. uzmandan geçme-kalma sınırı olarak ifade edilebilecek minimum yeterlik düzeyindeki 100 öğrenciden kaç öğrencinin 26 maddeden oluşan testin 1. maddesini doğru cevaplayabileceğini tahmin etmesi istenmiştir. Daha sonra aynı uzmanın testin diğer maddeleri için de aynı şekilde bir tahminde bulunması istenmiştir. Bulunan tüm değerler yüzde olarak değerlendirilip toplandıktan sonra madde sayısı olan 26'ya bölünerek 1. uzman için geçerli olan minimum geçme puanı (MGP) tespit edilmiştir. Aynı süreç diğer uzmanlar için de uygulanmış, her bir uzmandan elde edilen MGP'lerin ortalaması alınarak Angoff yöntemi için kesme puanı belirlenmiştir.

Nedelsky yöntemi ile yapılan 2. değerlendirmede bu kez 1. uzmandan minimum yeterlik düzeyindeki bir öğrencinin 1. maddenin seçeneklerinden kaç tanesini bilerek eleyebileceğini tahmin etmesi istenmiştir. Verilen yanıt 4 ise öğrencinin maddeyi doğru cevaplama ihtimali %100, 3 ise %50, 2 ise %33, 1 ise %25 ve 0 ise %20 olarak belirlenmiştir. 2. madde ve diğer maddeler için de elde edilen yüzdeler toplanıp madde sayısı olan 26'ya bölünerek 1. uzman için MGP elde edilmiştir. Aynı süreç diğer uzmanlar için de işletilmiş, elde edilen MGP'lerin aritmetik ortalaması alınarak Nedelsky yöntemi için kesme puanı belirlenmiştir.

Ebel yöntemi ile yapılan 3. değerlendirmede diğer iki yöntemden farklı olarak uzmanlardan her bir test maddesini “zorluk” ve “uygunluk” olarak adlandırılan, “zorluk”ta kolay-orta-zor ve “uygunlukta”ta gerekli-önemli-kabul edilebilir-tartışılabilir olmak üzere 3x4'lük çapraz bir tabloya yerleştirmeleri istenmiştir. Her madde tabloya yerleştirildikten sonra 1. uzmandan her hücrede bulunan soru ya da soruları minimum yeterlik düzeyindeki 100 öğrenciden kaç öğrencinin cevaplayabileceğini tahmin etmesi istenmiştir. Bulunan değer yüzdeye çevrilmiş, her hücredeki soru sayısı ile tahmin yüzdeleri çarpıldıktan sonra tüm değerler toplanmış ve madde sayısı olan 26'ya bölünerek 1. uzmanın MGP'si bulunmuştur. Süreç diğer uzmanlar için de aynı şekilde işletilmiş ve bulunan MGP'lerin aritmetik ortalaması alınarak Ebel yöntemi için kesme puanı belirlenmiştir.

3.4. Verilerin Analizi

Araştırmanın her bir alt probleminin çözümlenmesinde kullanılan istatistiksel yöntemler aşağıda verilmiştir.

Araştırmanın birinci alt problemi olan, üç farklı standart belirleme yöntemine göre belirlenen kesme puanlarının üzerinde puan alarak, başarılı kabul edilen öğrenci yüzdeleri arasında anlamlı bir farklılığın olup olmadığının araştırılmasında, bağımlı iki yüzde arasındaki farkın testi yapılmıştır. Bağımlı iki yüzde arasındaki farkın testi, yöntemlere göre başarılı ve başarısız sayılan öğrenci sayılarına ve yüzdelere ait 2x2'lik bir tablo yardımıyla hesaplanabilmektedir. Aşağıda Tablo 3-5'te bu 2x2'lik tabloya yer verilmiştir.

Tablo 3-5. Bağımlı gruplarda iki oran için genel tablo

		Yöntem 2		
		Başarılı	Başarısız	Toplam
Yöntem 1	Başarılı	a	b	a+b
	Başarısız	c	d	c+d
	Toplam	a+c	b+d	a+b+c+d=n

Tablo 3-5'te verilen a her iki yöntemden de başarılı olan kişi sayısını, d her iki yöntemden de başarısız olan kişi sayısını verirken, b yöntem 1'den başarılı olup yöntem 2'den başarısız olan kişi sayısını, c ise yöntem 1'den başarısız olup yöntem 2'den başarılı olan kişi sayısını göstermektedir.

Aynı bireylerden iki değişik zaman ya da iki değişik durumda elde edilen iki yüzde arasında fark olup olmadığını araştırmada bağımlı iki yüzde arasındaki farkın önemlilik testi uygulanır. Bağımlı örneklerde iki yüzde arasındaki farkın anlamlılık testinde z istatistiğinden faydalanılır. Tablo 3-5 yardımı ile karşılaştırılmak istenen iki yüzde; $p_1=(a+b)/n$ ve $p_2=(a+c)/n$ 'dir. Bu iki yüzde arasındaki farkın anlamlı olup olmadığına aşağıdaki eşitlik yardımı ile bakılabilir (Alpar, 2010)

$$z = \frac{b - c}{\sqrt{b + c}}$$

Gözlem sayısının az olması durumunda aşağıda verilen düzeltilmiş z istatistiğinden de yararlanılabilir.

$$z = \frac{|b - c| - 1}{\sqrt{b + c}}$$

Araştırmanın 1. alt probleminin çözümlenmesi sürecinde, 245 öğrencinin matematik başarı testi maddelerine verdikleri yanıtlar; doğrular 1, yanlışlar 0 olmak üzere Excel programına girilmiş, her öğrencinin toplam puanı madde sayısı olan 26'ya bölünerek başarı yüzdesi bulunmuştur. Üç farklı standart belirleme yönteminden elde edilen kesme puanlarına göre her bir yöntem için kesme puanının üzerinde kalanlar başarılı olmak üzere başarılı sayılan öğrenci sayısı ve yüzdesi hesaplanmıştır. Yöntemler ikiyeşerli gruplar halinde karşılaştırılarak farklı üç z istatistiği elde edilmiş, her bir yöntemle göre başarılı kabul edilen öğrenci yüzdeleri arasında anlamlı bir farklılığın olup olmadığına bakılmıştır.

Yöntemler arasında öğrencilerin başarılı-başarısız olarak sınıflandırılma durumları bakımından uyumun araştırıldığı araştırmanın 2. alt probleminin çözümlenmesinde Cohen'in Kappa Katsayısı kullanılmıştır. Kappa katsayısı $r=c$ (r sıra, c sütun) olan simetrik çapraz tablolarından hesaplanan bir orandır. Kappa katsayısı ile bir olayla ilgili, iki bağımsız gözlemci arasındaki değerlendirmelerin uyumu belirlenir. Kappa katsayısı, iki veya daha fazla gözlemci arasındaki uyuşmayı belirlemek için kategorik (nominal) nitelikteki verilerle kullanılır (Özdamar, 2004; Şencan, 2005). Kappa katsayısı -1 ile +1 arasında değerler alır. -1 negatif tam uyumu, 0 uyumsuzluğu, +1 ise pozitif tam uyumu gösterir (Karagöz, 2010). Kappa katsayısı 0,40 ile 0,75 arasında değer aldığı makul bir uyuşmadan, 0,75'ten büyük bir değer aldığı ise mükemmel bir uyuşmadan söz edilebilir (Şencan, 2005).

Yöntemler arasında uzmanlardan elde edilen minimum geçme puanları (MGP) bakımından uyumun araştırıldığı araştırmanın 3. alt probleminin çözümlenmesinde Pearson Momentler Çarpımı Korelasyon Katsayısı ve Bağımlı Gruplar t Testi kullanılmıştır. 19 uzmandan elde edilen MGP'ler arasındaki ilişki Pearson Momentler

Çarpımı Korelasyon Katsayısı ile yöntemlerden elde edilen kesme puanlarının ortalamaları arasındaki fark ise Bağımlı Gruplar t Testi ile incelenmiştir.

BÖLÜM IV

4. Bulgular ve Yorum

Bu bölümde, araştırma için toplanan veriler 2. bölümde kullanılan istatistiksel yöntemler kullanılarak analiz edilmiş, her bir alt probleme ait bulgulara ve bu bulgulara dair yorumlara yer verilmiştir.

4.1. Birinci Alt Probleme Ait Bulgular ve Yorumlar

Üç farklı standart belirleme yöntemine göre belirlenen kesme puanlarının üzerinde puan alarak, başarılı kabul edilen öğrenci yüzdeleri arasında anlamlı bir farklılık var mıdır?

Birinci alt problemin çözümü için 26 maddeden oluşan matematik başarı testinin her bir maddesi için, farklı üç standart belirleme yöntemi ile kesme puanı belirleme amacıyla, 19 uzmanın görüşlerine başvurulmuştur. Uzmanların Angoff, Nedelsky ve Ebel yöntemleri için vermiş oldukları bu kararlar Ek 3, Ek 4 ve Ek 5'te verilmiştir. Her uzmanın her bir yöntem için verdiği kararlar toplanıp aritmetik ortalaması alınarak her uzmanın yöntemlere göre MGP'si hesaplanmıştır. Bulunan MGP'lerin de aritmetik ortalaması alınarak farklı üç yöntemle göre farklı üç adet kesme puanı matematik başarı testi için hesaplanmıştır. Üç yöntem için de hesaplanan uzmanlara ait MGP'ler ve kesme puanları Tablo 4-1'de verilmiştir.

Tablo 4-1. Yöntemler için bulunan MGP'ler ve kesme puanları

Uzmanlar	Angoff	Nedelsky	Ebel
U1	64,80	27,84	67,50
U2	63,46	33,23	64,03
U3	56,15	40,88	47,69

Uzmanlar	Angoff	Nedelsky	Ebel
U4	30,76	26,30	30,96
U5	55,38	23,92	55,38
U6	50,00	37,92	57,50
U7	49,03	45,23	49,42
U8	51,26	43,73	50,76
U9	29,03	43,38	25,19
U10	42,69	31,96	40,76
U11	62,69	30,26	57,30
U12	31,53	27,65	25,00
U13	63,46	58,88	58,26
U14	22,30	41,03	21,92
U15	21,53	39,88	22,88
U16	31,46	39,19	27,11
U17	70,38	54,42	62,30
U18	44,42	34,15	48,46
U19	41,15	36,73	40,57
Kesme puanı	46,39	37,71	44,89

Tablo 4-1 incelendiğinde matematik başarı testi için en yüksek kesme puanının 46,39 değeriyle Angoff yöntemi tarafından elde edildiği görülmektedir. Angoff yöntemini 44,89 değeriyle Ebel yöntemi takip ederken, en düşük kesme puanının Nedelsky yöntemi tarafından 37,71 olarak belirlendiği görülmüştür. Bu bulgu Tanrıverdi (2006) ve Taşdemir (2013)'in Angoff ve Nedelsky yöntemlerini karşılaştırdığı ve Gündeğer (2012)'in Angoff ve Ebel yöntemlerini karşılaştırdığı çalışmaların sonuçlarıyla örtüşmektedir. Ancak, Taşdelen (2009)'in Nedelsky yönteminin belirlediği kesme puanının Angoff yönteminin belirlediği kesme puanından daha yüksek çıkan çalışmasının sonucu ile bu bulgu örtüşmemektedir.

Nedelsky yöntemi ile elde edilen kesme puanının diğer yöntemlerden daha düşük çıkmasının nedeni, uzmanların test maddelerinin seçenekleri içerisinde elenebilecek seçenek bulmakta zorlanmalarından kaynaklanabilir. Angoff yöntemi ile

elde edilen kesme puanının en yüksek çıkması, standart belirleme sürecinin diğerlerine oranla daha kolay olmasından kaynaklanabilir. Ebel yöntemiyle elde edilen kesme puanının Angoff yöntemiyle elde edilen kesme puanından daha düşük çıkmasının nedeni; Ebel yönteminin Angoff yöntemine göre biraz daha karmaşık olması ve puanlamasının daha zor olması olabilir. Ebel ve Angoff yöntemlerinin puanlarının nisbeten yakın çıkmasının nedeni; her ikisinde de uzmanlar tarafından geçme-kalma sınırındaki 100 öğrencinin düşünülüyor olması ve uzmanların zihinlerinde her iki yöntemde de maddeleri zorluk düzeyine göre kodluyor olmaları olabilir.

Her uzmanın test maddeleri hakkında verdikleri kararların farklılığı uzmanlara ait olan MGP'lerin de farklı olmasına neden olmuştur. Bu farklılık MGP'lerin ortalamasından elde edilen kesme puanlarının da her bir yöntem için farklı çıkması sonucunu beraberinde getirmiştir. Kesme puanının üzerinde puan alan öğrenciler başarılı kabul edildiğinden, kesme puanlarındaki bu farklılıklar başarılı kabul edilen öğrenci sayılarını ve öğrencilerin toplam öğrenci sayısı içindeki yüzdelerinin de farklı olmasına sebep olmuştur. Her bir yöntemin belirlediği kesme puanına göre başarılı kabul edilen öğrenci sayısı ve yüzdeleri Tablo 4-2'de gösterilmektedir.

Tablo 4-2. Farklı yöntemlerle belirlenen kesme puanları ve bu kesme puanlarına göre başarılı kabul edilen öğrenci sayısı ve yüzdeleri

Yöntemler	Kesme puanı	Başarılı öğrenci sayısı	Başarılı öğrenci yüzdesi
Angoff	46,39	107	0,43
Nedelsky	37,71	141	0,57
Ebel	44,89	117	0,47

Tablo 4-2'de görüldüğü gibi yöntemlerin belirlediği farklı kesme puanları başarılı kabul edilen öğrenci sayısının ve yüzdesinin de farklı olması sonucunu doğurmuştur. Kesme puanının en yüksek (46,39) olduğu Angoff yönteminde başarılı öğrenci sayısı (107) ve yüzdesi (0,43) en düşük olurken, kesme puanının en düşük (37,71) olduğu Nedelsky yönteminde başarılı kabul edilen öğrenci sayısı (141) ve yüzdesi (0,57) en yüksek olmuştur. Her bir yönteme göre başarılı kabul edilen öğrenci yüzdeleri arasındaki farklılıkların anlamlılığını bağımlı iki yüzde arasındaki farkın testi

ile sınamak için yöntemlere göre başarılı ve başarısız kabul edilen öğrencilerin sayı ve yüzdelerinin karşılaştırıldığı çapraz tablolar aşağıda gösterilmiştir.

Tablo 4-3. Angoff ve Nedelsky yöntemlerine göre başarılı-başarısız öğrenci sayısını gösteren çapraz tablo

Yöntemler		Nedelsky		
		Başarılı	Başarısız	Toplam
Angoff	Başarılı	107	0	107
	Başarısız	34	104	138
	Toplam	141	104	245

Tablo 4.3'te görüldüğü gibi Angoff yöntemine göre başarılı olup Nedelsky yöntemine göre başarısız öğrenci sayısı 0 iken, Nedelsky yöntemine göre başarılı olup Angoff yöntemine göre başarısız olan öğrenci sayısı 34'dür. Her iki yönteme göre başarılı olan öğrenci sayısı 107 iken, her iki yönteme göre başarısız öğrenci sayısı 104'dür.

Tablo 4-4. Angoff ve Ebel yöntemlerine göre başarılı-başarısız öğrenci sayısını gösteren çapraz tablo

Yöntemler		Ebel		
		Başarılı	Başarısız	
Angoff	Başarılı	107	0	107
	Başarısız	10	128	138
		117	128	245

Tablo 4-4'te görüldüğü gibi Angoff yöntemine göre başarılı olup Ebel yöntemine göre başarısız öğrenci sayısı 0 iken, Ebel yöntemine göre başarılı olup Angoff yöntemine göre başarısız olan öğrenci sayısı 10'dur. Her iki yönteme göre başarılı olan öğrenci sayısı 107 iken, her iki yönteme göre de başarısız olan öğrenci sayısı 128'dir.

Tablo 4-5. Nedelsky ve Ebel yöntemlerine göre başarılı-başarısız öğrenci sayısını gösteren çapraz tablo

Yöntemler		Ebel		
		Başarılı	Başarısız	
Nedelsky	Başarılı	117	24	141
	Başarısız	0	104	104
		117	128	245

Tablo 4-5'te görüldüğü gibi Nedelsky yöntemine göre başarılı olup Ebel yöntemine göre başarısız öğrenci sayısı 24 iken, Ebel yöntemine göre başarılı olup Angoff yöntemine göre başarısız olan öğrenci sayısı 0'dır. Her iki yönteme göre başarılı olan öğrenci sayısı 117 iken, her iki yönteme göre de başarısız olan öğrenci sayısı 128'dir. Her bir yönteme göre başarılı kabul edilen öğrenci yüzdeleri arasındaki farklılıkların anlamlılığı bağımlı iki yüzde arasındaki farkın testi ile sınanmış olup sonuçları Tablo 4-6'da verilmiştir.

Tablo 4-6. Yöntemlerin belirlediği kesme puanları, başarılı öğrenci sayı-yüzdesi ve z değerleri

Yöntemler	Kesme puanı	Başarılı öğrenci sayısı	Başarılı öğrenci yüzdesi	z
Angoff	46,39	107	0,43	-5,83**
Nedelsky	37,71	141	0,57	
Angoff	46,39	107	0,43	-3,16*
Ebel	44,89	117	0,47	
Nedelsky	37,71	141	0,57	4,89**
Ebel	44,89	117	0,47	

*p<.01; **p<.001

Tablo 4-6 incelendiğinde yöntemlerin belirlediği kesme puanlarına göre başarılı kabul edilen öğrenci yüzdeleri ikili olarak karşılaştırıldığında tüm yüzdeler arasında 0,05 düzeyinde manidar bir farklılığın olduğu görülmektedir. Bu sonuç

Tanrıverdi (2006) ve Taşdemir (2013)'in Angoff ve Nedelsky yöntemlerini karşılaştırdığı ve Gündeğer (2012)'in Angoff ve Ebel yöntemlerini karşılaştırdığı çalışmaların, başarılı kabul edilen öğrenci yüzdeleri arasında manidar bir farklılığın var olduğu sonuçlarıyla örtüşmektedir.

Başarılı kabul edilen öğrenci yüzdeleri arasındaki bu manidar farklılığın nedeni her bir yöntemin işleyiş sürecindeki farklılıkların kesme puanlarına yansımından kaynaklanıyor olabilir. Kesme puanlarındaki bu farklıklar doğal olarak başarılı kabul edilen öğrenci yüzdesini etkilemektedir. Her üç yöntemde de uzmanlar tarafından verilen kararlar farklı temellere dayanır. Angoff yönteminde uzmanlardan geçme-kalma sınırındaki 100 öğrenciden ilgili test maddesine kaç öğrencinin doğru yanıt verebileceğini tahmin etmeleri istenirken, Nedelsky yönteminde uzmanlardan geçme-kalma sınırındaki bir öğrencinin ilgili test maddesinin seçeneklerinden kaç tanesini eleyebileceğine odaklanması istenir. Ebel yönteminde ise diğer iki yöntemden çok daha farklı olarak uzmanlardan test maddelerinin her birini zorluk (orta-kolay-zor) ve uygunluk (önemli-gerekli-kabul edilebilir-tartışılabilir) olmak üzere 3x4 tipi bir tabloya yerleştirdikten sonra her hücredeki maddeleri geçme-kalma sınırındaki 100 öğrenciden kaç öğrencinin doğru yapabileceğini tahmin etmeleri istenir. Yöntemlerin işleyişindeki bu farklılıklar farklı MGP'lerin ve dolayısıyla farklı kesme puanların belirlenmesine neden olmuş, başarılı kabul edilen öğrenci yüzdeleri farklılaşmıştır.

4.2. İkinci Alt Probleme Ait Bulgular ve Yorumlar

Yöntemler arasında öğrencilerin başarılı-başarısız olarak sınıflandırılma durumları bakımından uyum var mıdır?

İkinci alt problemin çözümlenmesinde, öğrencilerin başarılı-başarısız olarak sınıflandırılmaları bakımından farklı standart belirleme yöntemleri arasındaki uyumu belirlemek amacıyla Cohen'in Kappa testi uygulanmıştır. İlişki düzeyinin incelenmesinde değişkenlerin yapısı dikkate alınmıştır. Öğrenci puanları normalde sürekli iken kesme puanının üzerinde kalanlar 1, altında kalanlar 0 olarak kodlanarak suni şekilde iki kesikli hale dönüştürülmüş ve Kappa testi uygulanmıştır. Kappa testi kategorik verilerin uyumunu ölçen bir testtir. Kappa katsayısı iki yöntemin tutarlığını

kontrol etmek için kullanılan, güvenilirlik hakkında bilgi veren bir katsayıdır. Yöntemler arasındaki uyumu belirlemek için yapılan Kappa testinin sonuçları Tablo 4-7’de verilmiştir.

Tablo 4-7. Yöntemler arasındaki uyumu gösteren kappa katsayıları

Yöntemler	Kappa	p
Angoff-Nedelsky	0,728	.00**
Angoff-Ebel	0,918	.00**
Ebel-Nedelsky	0,805	.00**

**p<.001

Kappa katsayısı 0,40 ile 0,75 arasında değer aldığımda makul bir uyuşmadan, 0,75’ten büyük bir değer aldığımda ise mükemmel bir uyuşmadan söz edilebilir (Şencan, 2005). Tablo 4-7 incelendiğinde tüm yöntemler arasında 0,01 düzeyinde uyum olduğu görülmektedir. Öğrencilerin başarılı-başarısız olarak sınıflandırılmaları bakımından, Angoff-Ebel yöntemleri (Kappa>0,75) ile Ebel-Nedelsky yöntemleri (Kappa>0,75) arasında yüksek derecede pozitif yönlü bir uyum olduğu görünürken, Angoff-Nedelsky yöntemleri (Kappa>0,40) arasında pozitif yönlü makul bir uyum olduğu görülmektedir.

Angoff ve Ebel yöntemi arasındaki yüksek uyumun nedeni uzmanların her iki yöntemde geçme-kalma sınırındaki 100 kişiye odaklanarak karar vermeleri olabilir. Nedelsky yönteminde diğer iki yöntemden farklı olarak geçme-kalma sınırındaki bir kişinin maddenin seçeneklerinden kaç tanesini eleyebileceğine odaklanılır. Angoff ve Ebel yöntemlerinin her ikisinde de uzmanlar maddeleri zihinlerinde kolay ve zor arası şeklinde kodlayarak öğrencilerin maddeleri yapabilirlikleri hakkında karar verirken, Nedelsky yönteminde uzmanlar soruların kolay ve zorluğundan ziyade seçeneklerin elenebilirliğine odaklanmaktadırlar. Angoff ve Nedelsky yöntemleri arasındaki kısmen daha düşük olan uyumun nedeni, yöntemlerin süreçlerinin birbirinden farklı olması olabilir. Nedelsky ve Ebel yöntemlerinin de süreçlerinin farklı olmasına rağmen uyumun yüksek olması dikkat çekici bir bulgudur. Bu bulgu uzman kararlarının birbirleriyle tutarlı olduğu ve yöntemlerin işleyişini doğru algıladıkları sonucunu doğurabilir.

4.3. Üçüncü Alt Probleme Ait Bulgular ve Yorumlar

Yöntemler arasında uzmanlardan elde edilen minimum geçme puanları (MGP) bakımından uyum var mıdır?

Üçüncü alt problemin çözümlenmesi amacıyla öncelikle her bir yöntem için 19 uzmandan elde edilen MGP'lerin betimsel istatistiklerine bakılmıştır. MGP'lere ait betimsel istatistikler ve normallik testi sonuçları Tablo 4-8'de verilmiştir.

Tablo 4-8. MGP'lere ait betimsel istatistikler ve normallik testi sonuçları

Betimsel İstatistikler	Angoff yöntemi	Nedelsky yöntemi	Ebel yöntemi
N	19	19	19
Ortalama	46,39	37,71	44,89
Ortanca	49,03	37,92	48,46
Mod	63,46	23,92 ^a	21,92 ^a
Standart sapma	15,30	9,19	15,27
Varyans	234,15	84,52	233,34
Çarpıklık katsayısı	-,157	,656	-,264
Basıklık katsayısı	-1,208	,359	-1,377
Shapiro Wilk Sig.	,298	,476	,087

a: Çoklu mod vardır ve en küçük olan gösterilmektedir.

Tablo 4-8 incelendiğinde her bir yönetime göre belirlenen MGP'lere ait ortalama, ortanca ve mod değerlerinin birbirine yakın değerler olduğu görülmektedir. Çarpıklık ve basıklık değerlerinin 0 değerinden çok sapmadığı, genel itibari ile -1 ve +1 arasında değiştiği gözlemlenmektedir. Shapiro Wilk testi, veri sayısının 50'den az olduğu durumlarda kullanılan bir normallik testidir (Büyüköztürk, 2008, s. 42). Shapiro Wilk Significant değeri 0,05'den büyük olduğunda dağılımın normal olduğu söylenebilmektedir. Shapiro Wilk testi p değeri her yöntem için 0,05 değerinden büyük çıkmıştır. Bu bilgiler ışığında yöntemlere ait MGP'lerin normal dağılımdan anlamlı bir fark göstermediği söylenebilir.

MGP'ler açısından yöntemler arasındaki uyumun araştırılması için MGP'lerin yapısı da dikkate alınarak Pearson Momentler Çarpımı Korelasyon Katsayısı kullanılmıştır. Yöntemlere ait MGP'ler arasındaki ilişkiyi gösteren sonuçlar Tablo 4-9'da verilmiştir.

Tablo 4-9. Uzmanlardan elde edilen MGP'ler arası pearson momentler çarpımı korelasyon katsayısı sonuçları

Yöntemler	Pearson Korelasyon Katsayısı	p	N
Angoff-Nedelsky	0,176	0,471	19
Angoff-Ebel	0,963	0,00*	19
Nedelsky-Ebel	0,088	0,722	19

*p<.001

Tablo 4-9 incelendiğinde, Angoff ve Nedelsky yöntemleri arasında uzmanlardan elde edilen MGP'ler bakımından düşük düzeyde bir ilişki olduğu görülmektedir. Bu bulgu Taşdemir (2013)'in MGP'ler arasında bulduğu ilişki bakımından benzerlik göstermemektedir. Düşük ilişkinin nedeni Nedelsky yönteminde elenebilecek seçenek sayısını kestirmede yaşadıkları tereddüt ve bazı uzmanların elenebilecek seçenek sayısı yerine elenemeyecek seçenek sayısını dikkate almaları olabilir.

Angoff ve Ebel yöntemleri arasında yüksek düzeyde anlamlı bir ilişki olduğu görülmektedir. Bu bulgu Gündeğer (2012)'in araştırma sonuçları ile örtüşmektedir. Yöntemler arasındaki bu yüksek ilişkinin nedeni her iki yöntemde de benzer şekilde geçme-kalma sınırındaki 100 öğrenciye odaklanması ve maddelerin güçlük indekslerinin dikkate alınması olabilir.

Nedelsky ve Ebel yöntemleri arasında düşük düzeyde bir ilişki olduğu görülmektedir. Bu bulgunun nedeni Nedelsky yönteminde geçme-kalma sınırındaki bir öğrencinin, Ebel yönteminde geçme-kalma sınırındaki 100 öğrencinin karar vermede dikkate alınıyor olması olabilir. Matematik sorularının seçeneklerden yapılabilme ihtimalinin diğer derslere göre daha zor olması nedeniyle, Nedelsky yöntemi ile standart belirleme sürecinde karar vermek uzmanlara zor gelmiş olabilir. Pearson Momentler Çarpımı Korelasyon Katsayısı'nın ortalamalardan bağımsız bir hesaplama yapması

nedeniyle yöntemlere ait MGP'lerin ortalamaları arasındaki farkın anlamlılığına bağlı gruplar t testi ile bakılmış, sonuçlar Tablo 4-10'da verilmiştir.

Tablo 4-10. MGP'lere ait ortalamalar arası farkın anlamlılığı için yapılan t testi sonuçları

Yöntemler	$\bar{X}$	t	sd	p
Angoff	46,39	2,306	18	0,033*
Nedelsky	37,71			
Angoff	46,39	1,568	18	0,134
Ebel	44,89			
Nedelsky	37,71	-1,828	18	0,084
Ebel	44,89			

*p<.01

Tablo 4-10 incelendiğinde Angoff ve Nedelsky yöntemlerine ait MGP'lerin ortalamasının bir başka ifade ile kesme puanlarının sırası ile 46,39 ve 37,71 olduğu görülmektedir. Yapılan t testi sonuçlarına göre bu iki yönteme ait ortalamalar arasındaki fark anlamlıdır [$t_{(18)}=0,033$, $p<.05$]. Bir başka ifade ile Angoff yöntemi ile belirlenen kesme puanı Nedelsky yöntemi ile belirlenen kesme puanından anlamlı şekilde daha büyüktür. Bu bulgu Taşdemir (2013)'in çalışmasının sonucu ile örtüşmektedir.

Tablo 4-10'da Angoff ve Ebel yöntemine ait kesme puanlarının sırası ile 46,39 ve 44,89 olduğu görülmektedir. T testi sonuçlarına göre bu iki yönteme ait MGP'in ortalamaları arasındaki fark anlamlı değildir [$t_{(18)}=1,568$, $p>.05$]. Bir başka deyişle Angoff yöntemi ile Ebel yönteminin belirlediği kesme puanları arasındaki yaklaşık 1,5 puanlık fark anlamlı bir fark değildir. Bu bulgu Gündeğer (2012)'in araştırmasının sonucu ile örtüşmemektedir. Gündeğer çalışmasını 17 uzman ve 489 öğrenci kullanarak gerçekleştirmiş, araştırmasında 20 maddelik matematik başarı testi kullanmış, başarı testini 4. sınıf öğrencilerine uygulamıştır. Uzman sayısındaki benzerliklere rağmen bulgulardaki bu farklılık, kullanılan öğrenci sayısı ve öğrencilerin sınıf düzeyinden kaynaklanıyor olabilir.

Tablo 4-10’da Nedelsky ve Ebel yöntemlerine ait kesme puanlarının sırası 37,71 ve 44,89 olduğu görülmektedir. Yapılan t testi sonuçlarına göre bu iki yönteme ait MGP’lerin ortalamaları arasındaki fark anlamlı değildir [$t_{(18)}=-1,828$, $p>.05$]. Bir başka ifade ile Ebel ve Nedelsky yöntemlerinin belirlediği yaklaşık 7 puanlık fark anlamlı değildir.

Bu bulgular sonucunda, Angoff ve Nedelsky yöntemleri ile elde edilen MGP’ler arasında anlamlı bir ilişki olmadığı, her iki yöntemden elde edilen MGP’lerin ortalamaları arasında ise manidar bir farklılığın olduğu görülmektedir. Bu anlamda Angoff ve Nedelsky yöntemleri arasında MGP’ler bakımından uyum olmadığı ve Angoff yöntemi ile belirlenen kesme puanının daha yüksek olduğu söylenebilir.

Angoff ve Ebel yöntemleri ile elde edilen MGP’ler arasında pozitif ve yüksek düzeyde bir ilişki söz konusu iken, bu iki yönteme göre belirlenen MGP’lerin ortalamaları arasında manidar bir farklılığın olmadığı görülmektedir. Bu bulgu iki yönteme ait MGP’ler bakımından yöntemler arasında uyumun var olduğunu ve Angoff yönteminden elde edilen kesme puanının Ebel yönteminden elde edilen kesme puanından anlamlı şekilde büyük olmadığını göstermektedir.

Nedelsky ve Ebel yöntemleri ile elde edilen MGP’ler arasında anlamlı bir ilişkinin varlığı söz konusu değilken, aynı şekilde bu iki yönteme ait MGP’lerin ortalamaları arasında da anlamlı bir farklılık mevcut değildir. Bu bulgu iki yönteme ait MGP’ler arasındaki ilişki bakımından yöntemler arasında bir uyum olmadığını ancak MGP’lerin ortalamaları bakımından yöntemler arasında bir uyum olduğunu göstermektedir.

BÖLÜM V

5. Sonuç ve Öneriler

Bu bölümde, araştırmanın bulguları doğrultusunda elde edilen sonuçlar özetlenmiş, bu sonuçlara dayalı olarak yapılabilecek önerilere yer verilmiştir.

5.1. Sonuçlar

Bu araştırma ile farklı üç standart belirleme yöntemi ile belirlenen kesme puanları, kesme puanlarına göre başarılı kabul edilen öğrenci yüzdeleri ve yöntemlerden elde edilen MGP'ler karşılaştırılmıştır. Araştırmadan elde edilen sonuçlar alt problemlerin sırası da dikkate alınarak aşağıda özetlenmiştir.

- Araştırma ile incelenen farklı üç standart belirleme yönteminin birbirinden farklı kesme puanları belirlediği görülmektedir. En yüksek kesme puanını belirleyen yöntemin Angoff (46.39), onu takip eden yöntemin Ebel (44.89) ve en düşük kesme puanı belirleyen yöntemin Nedelsky (37.71) olduğu görülmüştür. Farklı geçme puanları belirleyen standart belirleme yöntemlerine göre kesme puanının üzerinde puan alarak başarılı kabul edilen öğrenci yüzdeleri de farklılaşmaktadır. Angoff yönteminde öğrencilerin % 43'ü başarılı kabul edilirken, Ebel yönteminde bu oran % 47 olmuş ve Nedelsky yönteminde % 57'ye kadar yükselmiştir. Her üç yöntemle belirlenen kesme puanlarına göre başarılı sayılan öğrenci yüzdeleri arasında anlamlı bir farklılığın olduğu sonucuna ulaşılmıştır.
- Her bir yönteme göre hesaplanan kesme puanlarına göre başarılı öğrenciler 1 başarısız öğrenciler 0 şeklinde kodlanarak yöntemler arasındaki uyum incelenmiş ve yöntemler arasında anlamlı düzeyde uyum olduğu görülmüştür. Öğrencilerin başarılı-başarısız olarak sınıflandırılmaları bakımından, Angoff-Ebel yöntemleri ile Ebel-Nedelsky yöntemleri arasında yüksek derecede pozitif

yönlü bir uyum olduğu görülürken, Angoff-Nedelsky yöntemleri arasında pozitif yönlü makul bir uyum olduğu görülmüştür.

- Uzman kararlarından elde edilen MGP'ler arası ilişki bakımından, Angoff ve Nedelsky yöntemleri arasında uyum olmadığı, MGP'lerin ortalamalarının anlamlı şekilde farklı olmasından dolayı MGP'lerin ortalamaları bakımından uyum olmadığı görülmüştür. Bu sonuç Angoff ve Nedelsky yöntemlerinin birbirine benzer sonuçlar vermediğini göstermektedir. Angoff ve Ebel yöntemleri arasında MGP'ler arası ilişki yüksek, MGP'lerin ortalamaları arasındaki fark ise anlamlı değildir. Bu durum Angoff ve Ebel yöntemleri arasında hem MGP'ler hem de MGP'lerin ortalamaları bakımından bir uyum olduğu sonucunu ortaya çıkarmıştır. Bir başka ifade ile Angoff ve Ebel yöntemlerinin benzer sonuçlar verdiği görülmektedir. Nedelsky ve Ebel yöntemleri arasında elde edilen MGP'ler arası ilişki bakımından uyum olmadığı ancak MGP'ler arasındaki farkın anlamlı olmamasından dolayı MGP'lerin ortalamaları bakımından uyum olduğu görülmüştür.

5.2. Öneriler

5.2.1. Araştırma sonuçlarından çıkarılan öneriler

- Araştırma sonucunda belirlenen kesme puanı en yüksek Angoff yöntemi ile, en düşük Nedelsky yöntemi ile belirlenmiştir ve her yöntem arasında başarılı kabul edilen öğrenci yüzdesi arasında anlamlı fark vardır. Bu doğrultuda uygulanan sınavın çeşidine göre uygulayıcılar tarafından daha çok öğrencinin başarılı olması isteniyorsa Nedelsky yöntemi, daha az öğrencinin başarılı olması isteniyorsa Angoff yönteminin tercih edilmesi önerilebilir.
- Sınava katılanları başarılı-başarısız kabul etme, MGP'ler arası ilişki ve MGP ortalamaları arasındaki fark gibi üç farklı kriter açısından Angoff ve Ebel yöntemleri arasında var olan uyum nedeniyle ikisinin aynı anda kullanılacağı durumda ekonomik olması bakımından birinin kullanılması standart belirleme işlemini yapacaklara önerilebilir.

- Her ne kadar Nedelsky yönteminin ne Angoff ile ne de Ebel yöntemi ile hem MGP'ler hem de MGP'lerin ortalamaları bakımından uyum göstermediği görülse de Nedelsky yönteminin Angoff yöntemine göre Ebel yöntemiyle daha uyumlu olduğu test belirleme sürecinde göz önünde bulundurulmalıdır.
- Nedelsky yöntemi ile belirlenen kesme puanının diğerlerine göre daha düşük çıkmasında matematik testinin etkili olduğu düşünülmektedir. Matematik'in yapısı gereği her zaman seçeneklerden madde doğru yanıtlanamayabileceği için, Matematik gibi bir ders için Nedelsky yönteminin kullanılmaması standart belirleme işlemini yapacaklara önerilebilir.
- Mili Eğitim Sınıf Geçme Yönetmeliği'nde yapılan değişiklikle 45 olan geçer not 2013-2014 Eğitim-Öğretim Yılı itibarıyla 50 yapılmıştır. Bu araştırma sonucunda ise 9. sınıflar matematik başarı testi için geçme notu en yüksek Angoff yöntemi ile olmak üzere 46.39 bulunmuştur. Bu yönde yapılacak bir dizi çalışmanın benzer sonuçlar vermesi durumunda bu bulgular karar vericiler ile paylaşılarak geçme notunun yeniden 45 olması yönünde fikir beyan edilebilir.

5.2.2 Araştırmacılara yönelik öneriler

- Bu çalışmada test merkezli standart belirleme yöntemlerinden Angoff, Nedelsky ve Ebel yöntemleri kullanılmıştır. Benzer karşılaştırmalar öğrenci merkezli standart belirleme yöntemleri ile ya da diğer test merkezli standart belirleme yöntemleri kullanılarak da yapılabilir.
- Standart belirleme sürecinde çoktan seçmeli matematik başarı testi kullanılmıştır. Benzer bir çalışma farklı soru tiplerinden oluşan bir test içinde yapılabilir.
- Standart belirleme sürecinde uzmanların yaşadıkları problemlerin tespit edilebilmesi adına, uzmanlarla görüşülerek nitel bir çalışma yapılabilir.
- Standart belirleme sürecinde karar veren uzmanların deneyim durumlarına göre verdikleri kararlar karşılaştırılabilir.
- Hangi standart belirleme yönteminin daha kullanışlı, kolay ve ekonomik olduğuna dair uzman görüşlerinin alındığı bir araştırma yapılabilir.

KAYNAKÇA

- Alpar, R. (2010). *Spor, sađlık ve eđitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenirlik..* Ankara: Detay Yayıncılık.
- Andrew, B. J. ve Hecht, J. T. (1976). A preliminary investigation of two procedures for setting examination standards. *Educational and Psychological Measurement*, 36(1), 45 – 50.
- Appleton, L. P. (1984). *A comparison of the nedelsky, ebel, and contrasting groups methods of cut score determination.* Doktora tezi, USA: The Catholic University of America.
- Atılgan, H. (2011). Test geliştirme. H. Atılgan (ed), *Eđitimde ölçme ve deđerlendirme* (ss. 315-348) İçinde. Ankara: Anı Yayıncılık.
- Baykul, Y. (2000). *Eđitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması.* Ankara: ÖSYM Yayınları.
- Berberođlu, G. (2009). Madde haritalama yöntemi ve Cito Türkiye öğrenci izleme sistemi (ÖİS) uygulamalarında yeterlik düzeylerinin belirlenmesi. *Cito Eđitim: Kuram ve Uygulama*, 9(3), 9-24.
- Berk, R. A. (1986). A consumer's guide to setting performance standards on criterion-referenced tests. *Review of Educational Research*, 56(1), 137-172.
- Brennan, R. L. ve Lockwood, R. E. (1980). A comparison of the nedelsky and angoff cutting score procedures using generalizability theory. *Applied Psychological Measurement*, 4(2), 219 – 240.

- Buckendahl, C. W., Smith, W. R., Impara, J. C. ve Plake, B. S, (2002). A comparison of angoff and bookmark standard setting methods. *Journal of Educational Measurement*, 39(3), 253-263.
- Büyüköztürk, Ş. (2008). *Sosyal bilimler için veri analizi el kitabı* (9. Baskı). Ankara: Pegem Akademi.
- Chang, L. (1996). *Comparison between the nedelsky and angoff standart-setting methods*. Paper presented at the Annual Meeting of the National Council on Measurement in Education. New York, USA.
- Chang, L. (1999). Judgmental item analysis of the nedelsky and angoff standard-setting methods. *Applied Measurement in Education*, 12(2), 151-165.
- Chang, L., Van Der Linden, W. J. ve Vos, H. J. (2004). Setting standards and detecting intrajudge inconsistency using interdependent evaluation of response alternatives. *Educational and Psychological Measurement*, 64(5), 781-801.
- Cizek, G. J. (2001). *Setting performance standards: Concepts, methods and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cizek, G. J. ve Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. California: Sage Publication.
- Clauser, B. E., Harik, P., Margolis, M. J., McManus, I. C., Mollon, J., Chis, L. ve Williams, S. (2009). An empirical examination of the impact of group discussion and examinee performance information on judgements made in the angoff standard setting procedure. *Applied Measurement in Education*, 22, 1-21.
- Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*. NY: CBS College Publishing Company.

- Çetin, S. (2011). *İşaretleme ve angoff standart belirleme yöntemlerinin karşılaştırılması*. Doktora Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Çetin, S. ve Gelbal, S. (2010). Farklı standart belirleme yöntemlerinin geçme puanları üzerine etkisi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 43(1), 79-95.
- Downing, S. M., Tekian, A. ve Yudkowsky, R. (2006). Research methodology: Procedures for establishing defensible absolute passing scores on performance examinations in health professions education. *Teaching and Learning in Medicine*, 18(1), 50-57.
- Ertürk, S. (1994). *Eğitimde program geliştirme* (8. Baskı). Ankara: Meteksan Yayınevi.
- Gündeğer, C. (2012). *Angoff, yes/no ve ebel standart belirleme yöntemlerinin karşılaştırılması*. Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Hess, B., Subhiyah, R. G. ve Giordano, C. (2007). Convergence between cluster analysis and the Angoff method for setting minimum passing scores on credentialing examinations. *Evaluation and the Health Professions*, 30(4), 362-375.
- Hambleton, R. K. (1997). Standard setting in criterion referenced tests. J. P. Keeves (ed.), *Educational research, methodology, and measurement: An international handbook* (ss. 798-803) İçinde. Great Britain, UK: Cambridge University Press.
- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. G. J. Cizek (ed.), *Setting performance standards: Concepts, methods and perspectives* (ss. 89-116) İçinde, Mahwah, NJ: Lawrence Erlbaum Associates.

- Harasym, P. H. (1981). A comparison of the nedelsky and modified angoff Standard setting procedure on evaluation outcome. *Educational and Psychological Measurement*, 41(3), 725-734.
- Hurtz G. M. ve Hertz, N. R. (1999). How many raters should be used for establishing cutoff scores with the angoff method? A generalizability theory study. *Educational And Psychological Measurement*, 59(6), 885 – 897.
- Huyhn, H. (2000). *On item mappings and statistical rules for selecting binary items for criterion-referenced interpretation and bookmark standard setting*. Paper presented Annual Meeting of National Council on Measurement in Education, New Orleans, LA.
- Impara, J. C. ve Plake, B. S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34(4), 353-366.
- Impara, J. C. ve Plake, B. S. (1998). Teachers' ability to estimate item difficulty: a test of the assumptions in the angoff standard setting method. *Journal of Educational Measurement*, 35(1), 69-81.
- Impara, J. C., Plake, B.S. ve Irwin, P. (2000). Consistency of angoff-based predictions of item performance: Evidence of technical quality of results from the angoff standard setting method. *Journal of Educational Measurement*, 37(4), 247-355.
- Irwin, P. (2007). *An alternative examinee-centered standard setting strategy*. Doktora tezi, USA: University of Nebraska.
- Jeager, R. M. (1989). Certification of student competence. R. L. Linn (ed), *Educational measurement* (ss. 485-514) İçinde. New York: American Council on Education/Macmillan.

- Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research*, 63(4), 425-461.
- Kane, M. T. (1998). Choosing between examinee-centered and test-centered standard-setting methods. *Educational Assessment*, 5(3), 129-145.
- Karagöz, Y. (2010). *İlişki katsayıları*. Ankara: Detay Yayıncılık.
- Karasar, N. (1982). *Bilimsel araştırma yöntemi*. Ankara: Matbaş Matbaacılık ve Ambalaj Sanayi.
- Kaptan, S. (1977). *Bilimsel araştırma teknikleri*. Ankara: Tekışık Matbaası ve Rehber Yayınevi.
- Kubiszyn, T. ve Borich G. (2007). *Educational testing and measurement classroom application and practice*. USA: Wiley Jossey-Bass Education.
- Livingston, S. A. ve Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests*. New Jersey: Educational Testing Service.
(http://www.ets.org/Media/Research/pdf/passing_scores.pdf, 10.01.2014 tarihinde erişildi).
- Livingston, S. A. ve Zieky, M. J. (1983). *A comparative study of standard-setting methods*. Princeton: Educational Testing Service.
- Livingston, S. A. ve Zieky J. M. (1989). A comparative study of standard-setting methods. *Applied Measuremet in Education*, 2(2), 121-141.
- Mills, C. N. (1983). A comparison of three methods of establishing cut-off scores on criterion-referenced tests. *Journal of Educational Measurement*, 20(3), 283-292.
- Mills, C. N. ve Berr, J. E. (1983). A comparison of standard setting methods: do the same judges establish the same standards with different methods? Paper

presented at the Annual Meeting of the American Educational Research Association: Montreal, Canada.

Mills, C. N. ve Melican, G. J. (1988). Estimating and adjusting cutoff scores: Features of selected methods. *Applied Measurement in Education*, 1(3), 261 – 275.

Mills, C. N., Melican, G. J. ve Thomas Ahluwaila, N. (1991). Defining minimal competence. *Educational Measurement: Issues and Practice*, 10(2), 7-10.

Mitzel, H. C., Lewis, D. M., Patz, R. J. ve Green, D. R. (2001). “The bookmark procedure: Psychological perspectives. G. J. Cizek (ed.), *Setting performance standards: Concepts, methods and perspectives* (ss. 53-88) İçinde. Mahwah, NJ: Lawrence Erlbaum Associates.

Murphy, K. R. ve Davidshofer, C. O. (1991). *Psychological testing: Principles and applications*. New Jersey: Prantice Hall.

National Research Council (2001). *Knowing what students know. The science and design of educational assesment*. National Academic Press. Washington, DC.

Nitko, A. J. ve Brookhart, S. M. (2007). *Educational assessment of students*. New Jersey: Pearson Education.

Norcini, J. J., Lipner, R. S., Langdon, L. O. ve Strecker, C. A. (1987). A comparison of three variations on a standard-setting method. *Journal of Educational Measurement*, 24(1), 56-64.

Osterlind, S. J. (1998). *Constructing test item*. Boston: Kluwer Academic Publihers.

Ömür, S. ve Selvi, H. (2010). Angoff, ebel ve nedelsky yöntemleriyle belirlenen kesme puanlarının sınıflama tutarlılıklarının karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 109-113.

- Öner, N. (1997). *Türkiye’de kullanılan psikolojik testler: Bir başvuru kaynağı*. İstanbul: Boğaziçi Üniversitesi Yayınları.
- Özcelik, D. A. (1992). *Ölçme ve değerlendirme*. Ankara: ÖSYM.
- Özdamar, K. (2004). *Paket programlar ile istatistiksel veri analizi (Çok değişkenli analizler)*. Eskişehir: Kaan Kitapevi.
- Poggio, J. P., Glasnapp, D. R. ve Eros, P. I. (1982). An Evaluation of Contrasting Groups Methods for Setting Standards. Paper presented at The Annual Meeting of the American Educational Research Association. New York: USA.
- Reckase, M.D. (2006). A conceptual framework for a psychometric theory for standard setting with examples of its use for evaluating the functioning of two standard setting methods. *Educational Measurement: Issues and Practice*, 25(2), 4-18.
- Shen, L. (2001). A comparison of angoff and rasch model based item map methods in standard setting. *Paper Presented at the Annual Meeting of the American Educational Research Association, Seattle, WA*.
- Schulz, M. E. (2006). Commentary: A response to reckase’s conceptual framework and examples for evaluating standard setting methods. *Educational Measurement: Issues and Practice*, 25(3), 4-13.
- Skakun, E. N. ve Kling, S. (1980). Comparability of methods for setting standards. *Journal of Educational Measurement*, 17(3), 229 – 235.
- Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlilik*. Ankara: Seçkin Yayıncılık.

- Tanrıverdi, S. (2006). *Standart belirleme yöntemlerinin geçme puanları üzerine etkisi*. Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Taşdelen, G. (2009). *Nedelsky ve angoff standart belirleme yöntemlerinin genellenebilirlik kuramı ile karşılaştırılmasına ilişkin bir araştırma*. Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Taşdemir, F. (2013). *Angoff (1-0), nedelsky ve sınır değerleri saptama yöntemleri ile bir testin sınıflama doğruluklarının incelenmesi*. Doktora Tezi, Ankara: Ankara Üniversitesi Eğitim Bilimleri Enstitüsü.
- Tekin, H. (1979). *Eğitimde ölçme ve değerlendirme*, Ankara: Mars Matbaası.
- Turgut, M. F. (1993). *Eğitimde ölçme ve değerlendirme metotları*. Ankara: Saydam Matbaacılık.
- Tülübaş, G. (2009). *Psikolojik testlerde angoff ve sınır grup yöntemleri ile kesme puanlarının belirlenmesi*. Yüksek Lisans Tezi, Ankara: Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü.
- Wang, L., Pan, W., ve Austin, J. T. (2003). Standards – setting procedures in accountability research: Impacts of conceptual frameworks and mapping procedures on passing rates. *Paper presented at the annual meeting of the American Educational Research Association, Chicago*.
- Zieky, M. ve Perie, M. (2004). *A primer on setting cut scores on tests of educational achievements*, ETS Publication.
- Zwick, R., Şentürk, D. ve Wang, J. (2001). An investigation of alternative methods for item mapping in the national assessment of educational progress. *Educational Measurement: Issues and Practice*, 20(2), 15-25.

EKLER

Ek 1. Matematik başarı testi

MATEMATİK BAŞARI TESTİ

ADINIZ.....
 SOYADINIZ.....
 OKULUNUZ.....
 SINIFINIZ.....
 9. SINIF 1. DÖNEM SONU MATEMATİK KARNE NOTUNUZ....

AÇIKLAMA:

Değerli öğrenciler,

Bu matematik başarı testinde 9. sınıf 1. dönem müfredatında bulunan kümeler ve denklemeşitsizlikler öğrenme alanlarından toplam 34 soru bulunmaktadır. Test sonunda elde edilen puanlar öğrenme düzeyinizi belirlemek amacıyla kullanılacaktır. Ayrıca puanlar, standart belirleme yöntemleri kullanılarak geçme puanı elde edildikten sonra test puanlarına göre öğrencilerin sınıflandırılmasını inceleyen yüksek lisans tezinde veri olarak değerlendirilecektir. Cevaplama süresi bir ders saatidir ve her sorunun bir doğru cevabı vardır. Araştırma sonuçları başka bir amaçla kullanılmayacaktır. Katılımlarınızdan dolayı teşekkür ederim.

Okan DEMİR
 AİBÜ Ölçme ve Değerlendirme ABD

Lütfen her soruyu yanıtlamaya özen gösteriniz.

SORULAR

1) Aşağıdakilerden hangisi bir küme belirtmez?

- A) Asal sayılar
- B) Türkiye'deki iller
- C) Haftanın a ile başlayan günleri
- D) Güzel harfler
- E) Okulunuzdaki öğretmenler

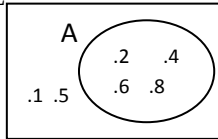
2) $A = \{x: 4 < x \leq 7, x \in \mathbb{N}\}$ kümesinin liste biçiminde gösterimi aşağıdakilerden hangisidir?

- A) {4,5,6,7} B) {5,6,7} C) {4,7} D) {5,6} E) {6,7}

3) Aşağıdakilerden hangisi boş küme değildir?

- A) B ile başlayan günler
 B) $K = \{x: x^2 + 1 = 0, x \in \mathbb{R}\}$
 C) $L = \{x: x, \text{gün sayısı } 25 \text{ olan ay}\}$
 D) $M = \{x: 7 < x < 8, x \in \mathbb{N}\}$
 E) $N = \{x: x, \text{Türkiyedeki boyu } 2 \text{ metre olan insanlar}\}$

4) E



Yandaki şekle göre
aşağıdakilerden kaç tanesi
doğrudur?

- I. 8, A'nın elemanıdır
 II. $4 \notin A$
 III. $s(A) = 4$
 IV. A kümesi sonlu bir kümedir

- A) 0 B) 1 C) 2 D) 3 E) 4

5) 5 elemanlı bir kümenin özalt küme sayısı kaçtır?

- A) 1 B) 3 C) 7 D) 15 E) 31

6) $A = \{a, b, c, d, e\}$ kümesinin alt kümelerinin kaç tanesinde a bulunur, b bulunmaz?

- A) 4 B) 8 C) 16 D) 32 E) 36

7) $A = \{x: 4 < x < 8, x \in \mathbb{N}\}$ kümesi aşağıdakilerden hangisine eşittir?

- A) $K = \{x: 4 < x < 8, x \in \mathbb{R}\}$
 B) $L = \{4, 5, 6, 7, 8\}$
 C) $M = \{4, 5, 6, 7\}$
 D) $N = \{x: 16 < x^2 < 64, x \in \mathbb{N}\}$
 E) $P = \{x: 12 < 3x < 21, x \in \mathbb{N}\}$

8) $K = \{x: x^2 - 4 = 0, x \in \mathbb{Z}\}$, $L = \{-2, 2\}$ ve $M = \{-2, 0, 2\}$ olarak veriliyor. Aşağıdakilerden hangisi doğrudur?

- A) K sonsuz bir kümedir.
 B) $M \subset L$
 C) $s(K) = 3$
 D) $K = M$
 E) $K = L$

9) $A = \{1, 2, 3, 4\}$ $B = \{3, 4, 5\}$ $C = \{4, 5, 6, 7, 8\}$ ise $A \cap (B \cup C)$ kümesini bulunuz.

- A) $\{3, 4, 5\}$ B) $\{3, 4\}$ C) $\{4, 5\}$ D) $\{5\}$ E) $\{4\}$

10) $s(A) = 2s(B)$, $s(A - B) = 7$, $A \cap B$ nin alt küme sayısı 32 ise $s(A \cup B) = ?$

- A) 13 B) 11 C) 10 D) 8 E) 5

11) $(2x,5)=(6,3y-1)$ ise $x+y=?$

- A) 9 B) 8 C) 6 D) 5 E) 4

12) $s[(A \times B) \cap (A \times C)] = 12$, $s(B \cap C) = 3$ ise $s(A) = ?$

- A) 3 B) 4 C) 6 D) 12 E) 8

13) Erkek ve kız öğrencilerin bulunduğu 45 kişilik bir sınıfta gözlüklü erkeklerin sayısı gözlüksüz kızların sayısının 4 katıdır. Sınıfta 14 erkek gözlüksüz ve 6 kız gözlüklü olduğuna göre erkeklerin sayısı kaçtır?

- A) 34 B) 32 C) 29 D) 20 E) 19

14) Bir grupta Almanca ve Fransızca dillerinden sadece birini bilen 15, en az birini bilen 19, en çok birini bilen 29 kişi olduğuna göre bu grup kaç kişiliktir?

- A) 30 B) 31 C) 33 D) 35 E) 40

15) Aşağıdakilerden hangisi yanlıştır?

- A) π irrasyonel sayıdır.
 B) $\sqrt{2}$ reel sayıdır.
 C) $2, \bar{6}$ rasyonel sayıdır.
 D) $\sqrt[3]{-6}$ reel sayıdır.
 E) $\sqrt[4]{-7}$ reel sayıdır.

16) Aşağıdakilerden hangisi yanlıştır?

- A) $N^+ \subset N$ B) $N \subset Z$ C) $Z \subset Q$ D) $Q \subset R$ E) $Q' \subset Q$

17) $x, y \in Z$, $-2 < x < 4$ ve $0 < y < 5$ ise $3x - 2y$ 'nin alabileceği en büyük tamsayı değeri nedir?

- A) 7 B) 6 C) 5 D) 4 E) 3

18) $a^2b < 0$, $bc > 0$, $ac < 0$ ise a, b ve c'nin işaretleri sırasıyla aşağıdakilerden hangisidir?

- A) +++ B) +-- C) --+ D) --- E) +- -

19) $x, y \in R$, $-1 < x < 4$ ve $2x + y - 1 = 0$ olduğuna göre y'nin değerler aldığı aralık aşağıdakilerden hangisidir?

- A) (-7, -3) B) (-3, 7] C) (-3, 7) D) (-7, 3) E) [-7, 3)

20) $-4 < x < -2$ ise x^2 hangi aralıkta değerler alır?

- A) $4 < x^2 < 16$ B) $0 < x^2 < 16$ C) $0 < x^2 < 4$ D) $-4 < x^2 < 16$ E) $2 < x^2 < 16$

21) $4x - 5 < 2x + 7$ eşitsizliğini sağlayan kaç farklı doğal sayı vardır?

- A) 7 B) 6 C) 5 D) 4 E) 3

22) $2(x-1)-(x+1)=5$ denkleminin çözüm kümesinin bulunuz.

- A) {8} B) {7} C) {6} D) {5} E) {4}

23) $|2x - 5| = 7$ denkleminin çözüm kümesini bulunuz.

- A) {-1,6} B) {-6,1} C) {-6,-1} D) {1,6} E) {6}

24) $a < b < c$ olmak üzere $|a - b| + |c - b| - |a - c|$ nin eđiti nedir?

- A) -2 B) -1 C) 0 D) 1 E) 2

25) $\left. \begin{array}{l} 2x-y=2 \\ x+2y=1 \end{array} \right\}$ denklem sisteminin çözüm kümesi nedir?

- A) (1,0) B) (1,2) C) (1,3) D) (0,2) E) (1,1)

26) $a.b=18$
 $a.c=12$
 $b.c=24$ ise $a.b.c=?$

- A) 72 B) 68 C) 44 D) 36 E) 24

27) $\frac{2^5+2^5+2^5+2^5}{2^3 \cdot 2^3} = ?$

- A) 4 B) 2 C) 1 D) $\frac{3}{2}$ E) $\frac{1}{2}$

28) $3^x + 3^{x-2} = 30$ ise $x = ?$

- A) 0 B) 1 C) 2 D) 3 E) 4

29) $\sqrt{14 + \sqrt{1 + \sqrt{9}}} = ?$

- A) 8 B) 5 C) 4 D) 3 E) 2

30) $\sqrt{0,04} + \sqrt[3]{0,008} = ?$

- A) 0,6 B) 0,4 C) 0,2 D) 0,1 E) 1

31) Ali 1994 yılında doğmuştur. 2015 yılında Ali'nin yaşının Barış'ın yaşına oranı $\frac{7}{6}$ olacağına göre Barış hangi yıl doğmuştur?

- A) 2000 B) 1998 C) 1997 D) 1995 E) 1993

32) 3 usta bir günde 96 m^2 duvar boyuyor ise aynı nitelikte 4 usta bir günde kaç m^2 duvar boyar?

- A) 120 B) 126 C) 128 D) 130 E) 144

33) “Bir sayının 3 katının 20 eksiği, sayının yarısının 3 fazlasına eşittir” probleminin çözümü için kurulması gereken denklem aşağıdakilerden hangisidir?

A) $3 \cdot 20 - x = \frac{x}{2} + 3$

B) $3x - 20 = \frac{x}{3} + 2$

C) $3(x - 20) = \frac{x}{2} + 3$

D) $3x - 20 = \frac{x}{2} + 3$

E) $3x - 20 = \frac{x+3}{2}$

34) %32 kar edilerek 660 liraya satılan bir mal 440 liraya satılırsa yüzde kaç zarar edilir?

A) 18

B) 15

C) 14

D) 12

E) 10

Ek 2. Uzman görüşleri formu

UZMAN GÖRÜŞLERİ FORMU

Değerli öğretmenler,

Yüksek lisans tez çalışmasında kullanılmak amacıyla geliştirilen, 26 sorudan oluşan 9. sınıf başarı testi için bir geçme (kesme) puanı hesaplanacaktır. Bu puan başarılı öğrenci ile başarısız öğrenciyi ayıran sınır puandır. Sizden istenen yapılacak üç farklı değerlendirme için bazı tahminlerde bulunmanızdır. Yapılacak üç farklı değerlendirme sonucunda üç geçme puanı hesaplanacaktır. Sorular kümeler, denklem ve eşitsizlik öğrenme alanlarını kapsamaktadır. Değerlendirmelerinizi içten ve objektif olarak yapmanızı rica eder, şimdiden katkılarınız için teşekkür ederim.

Okan DEMİR

AİBÜ Ölçme ve Değerlendirme ABD

Değerlendirme 1:

Angoff yöntemi olarak bilinen yöntemle kesme puanı hesaplayabilmek için aşağıda bulunan 26 sorunun her biri için şu soruya cevap veriniz ve kararınızı her sorunun solunda bulunan kutucuğa yazınız.

“Sizce matematik dersinde geçme-kalma sınırı ya da sınır grup olarak ifade edilen minimum yeterlik düzeyindeki 100 öğrenciden kaç öğrenci bu soruyu doğru yanıtlar?”

Değerlendirme 2:

Nedelsky yöntemi olarak bilinen yöntemle kesme puanı hesaplayabilmek için aşağıda bulunan 26 sorunun her biri için şu soruya cevap veriniz ve kararınızı her sorunun sağında bulunan kutucuğa yazınız.

“Sizce matematik dersinde geçme-kalma sınırı ya da sınır grup olarak ifade edilen minimum yeterlik düzeyindeki bir öğrenci bu sorunun yanıtlarından kaç tanesini bilerek eleyebilir?”

DEĞERLENDİRME 1		DEĞERLENDİRME 2
	1) Aşağıdakilerden hangisi bir küme belirtmez? A) Asal sayılar B) Türkiye'deki iller C) Haftanın a ile başlayan günleri D) Güzel harfler E) Okulunuzdaki öğretmenler	
	2) Aşağıdakilerden hangisi boş küme değildir? A) B ile başlayan günler B) $K = \{x: x^2 + 1 = 0, x \in \mathbb{R}\}$ C) $L = \{x: x, \text{gün sayısı } 25 \text{ olan ay}\}$ D) $M = \{x: 7 < x < 8, x \in \mathbb{N}\}$ E) $N = \{x: x, \text{Türkiyedeki boyu } 2 \text{ metre olan insanlar}\}$	
	3) 5 elemanlı bir kümenin özalt küme sayısı kaçtır? A) 1 B) 3 C) 7 D) 15 E) 31	
	4) $A = \{a, b, c, d, e\}$ kümesinin alt kümelerinin kaç tanesinde a bulunur, b bulunmaz? A) 4 B) 8 C) 16 D) 32 E) 36	
	5) $A = \{x: 4 < x < 8, x \in \mathbb{N}\}$ kümesi aşağıdakilerden hangisine eşittir? A) $K = \{x: 4 < x < 8, x \in \mathbb{R}\}$ B) $L = \{4, 5, 6, 7, 8\}$ C) $M = \{4, 5, 6, 7\}$ D) $N = \{x: 16 < x^2 < 64, x \in \mathbb{N}\}$ E) $P = \{x: 12 < 3x < 21, x \in \mathbb{N}\}$	
	6) $K = \{x: x^2 - 4 = 0, x \in \mathbb{Z}\}$, $L = \{-2, 2\}$ ve $M = \{-2, 0, 2\}$ olarak veriliyor. Aşağıdakilerden hangisi doğrudur? A) K sonsuz bir kümedir. B) $M \subset L$ C) $s(K) = 3$ D) $K = M$ E) $K = L$	
	7) $A = \{1, 2, 3, 4\}$ $B = \{3, 4, 5\}$ $C = \{4, 5, 6, 7, 8\}$ ise $A \cap (B \cup C)$ kümesini bulunuz. A) $\{3, 4, 5\}$ B) $\{3, 4\}$ C) $\{4, 5\}$ D) $\{5\}$ E) $\{4\}$	

	8) $(2x,5)=(6,3y-1)$ ise $x+y=?$ A) 9 B) 8 C) 6 D) 5 E) 4	
	9) $s[(A \times B) \cap (A \times C)]=12$, $s(B \cap C)=3$ ise $s(A)=?$ A) 3 B) 4 C) 6 D) 12 E) 8	
	10) Erkek ve kız öğrencilerin bulunduğu 45 kişilik bir sınıfta gözlüklü erkeklerin sayısı gözlüksüz kızların sayısının 4 katıdır. Sınıfta 14 erkek gözlüksüz ve 6 kız gözlüklü olduğuna göre erkeklerin sayısı kaçtır? A) 34 B) 32 C) 29 D) 20 E) 19	
	11) Bir grupta Almanca ve Fransızca dillerinden sadece birini bilen 15, en az birini bilen 19, en çok birini bilen 29 kişi olduğuna göre bu grup kaç kişiliktir? A) 30 B) 31 C) 33 D) 35 E) 40	
	12) Aşağıdakilerden hangisi yanlıştır? A) $N^+ \subset N$ B) $N \subset Z$ C) $Z \subset Q$ D) $Q \subset R$ E) $Q' \subset Q$	
	13) $x,y \in Z$, $-2 < x < 4$ ve $0 < y < 5$ ise $3x-2y$'nin alabileceği en büyük tamsayı değeri nedir? A) 7 B) 6 C) 5 D) 4 E) 3	
	14) $a^2b < 0$, $bc > 0$, $ac < 0$ ise a, b ve c'nin işaretleri sırasıyla aşağıdakilerden hangisidir? A) +++ B) +-- C) --+ D) --- E) -+-	
	15) $-4 < x < -2$ ise x^2 hangi aralıkta değerler alır? A) $4 < x^2 < 16$ B) $0 < x^2 < 16$ C) $0 < x^2 < 4$ D) $-4 < x^2 < 16$ E) $2 < x^2 < 16$	
	16) $4x-5 < 2x+7$ eşitsizliğini sağlayan kaç farklı doğal sayı vardır? A) 7 B) 6 C) 5 D) 4 E) 3	

	17) $2(x-1)-(x+1)=5$ denkleminin çözüm kümesinin bulunuz. A) {8} B){7} C){6} D){5} E){4}	
	18) $2x - 5 = 7$ denkleminin çözüm kümesini bulunuz. A) {-1,6} B){-6,1} C){-6,-1} D) {1,6} E) {6}	
	19) $a < b < c$ olmak üzere $a - b + c - b - a - c$ nin eđiti nedir? A) -2 B) -1 C) 0 D) 1 E) 2	
	20) $\left. \begin{array}{l} 2x-y=2 \\ x+2y=1 \end{array} \right\}$ denkleminin çözüm kümesi nedir? A) (1,0) B) (1,2) C) (1,3) D) (0,2) E) (1,1)	
	21) $\frac{2^5+2^5+2^5+2^5}{2^3 \cdot 2^3} = ?$ A) 4 B) 2 C) 1 D) $\frac{3}{2}$ E) $\frac{1}{2}$	
	22) $3^x + 3^{x-2} = 30$ ise $x = ?$ A) 0 B) 1 C) 2 D) 3 E) 4	
	23) $\sqrt{0,04} + \sqrt[3]{0,008} = ?$ A) 0,6 B) 0,4 C) 0,2 D) 0,1 E) 1	
	24) Ali 1994 yılında doğmuştur. 2015 yılında Ali'nin yaşının Barış'ın yaşına oranı $\frac{7}{6}$ olacağına göre Barış hangi yıl doğmuştur? A) 2000 B) 1998 C) 1997 D) 1995 E) 1993	

	25) “Bir sayının 3 katının 20 eksiği, sayının yarısının 3 fazlasına eşittir” probleminin çözümü için kurulması gereken denklem aşağıdakilerden hangisidir? A) $3 \cdot 20 - x = \frac{x}{2} + 3$ B) $3x - 20 = \frac{x}{3} + 2$ C) $3(x - 20) = \frac{x}{2} + 3$ D) $3x - 20 = \frac{x}{2} + 3$ E) $3x - 20 = \frac{x+3}{2}$	
	26) %32 kar edilerek 660 liraya satılan bir mal 440 liraya satılırsa yüzde kaç zarar edilir? A) 18 B) 15 C) 14 D) 12 E) 10	

Değerlendirme 3:

Ebel yöntemi olarak bilinen bu yöntemle geçme puanı hesaplamak için kullanacağımız bu değerlendirmeyi yapmadan önce lütfen tüm soruları tekrar gözden geçirin. Bu değerlendirme iki aşamadan oluşmaktadır. Birinci aşamada her soruyu zorluk derecesine göre kolay, orta ve zor olmak üzere 3, uygunluk derecesine göre gerekli, önemli, kabul edilebilir ve tartışılabilir olmak üzere 4 kategoriye ayırıp aşağıda verilen tabloda sorular kısmına yazınız. Soruları testteki diğer sorulara göre kıyaslayınız. Örneğin, bir soru testteki diğer sorulara göre kolaysa kolay, daha kabul edilebilirse kabul edilebilir hücrelerine yerleştiriniz. İkinci aşamada ise 12 hücrenin her biri için şu soruya cevap verip, kararınızı her hücrede bulunan yüzde kısmına yazınız.

“Sizce matematik dersinde geçme-kalma sınırı ya da sınır grup olarak ifade edilen minimum yeterlik düzeyindeki 100 öğrenciden kaç tanesi bu hücredeki soruları doğru yanıtlayabilir?”

DEĞERLENDİRME 3				
		Zorluk		
		Kolay	Orta	Zor
Uygunluk	Gerekli	Sorular: Yüzde:	Sorular: Yüzde:	Sorular: Yüzde:
	Önemli	Sorular: Yüzde:	Sorular: Yüzde:	Sorular: Yüzde:
	Kabul edilebilir	Sorular: Yüzde:	Sorular: Yüzde:	Sorular: Yüzde:
	Tartışılabilir	Sorular: Yüzde:	Sorular: Yüzde:	Sorular: Yüzde:

Not: Her hücre dolmak zorunda değildir, fakat tüm sorular hücrelere yerleştirilmelidir.

Katkılarınız için teşekkür ederiz

Ek 3. Angoff yöntemi ile uzmanlardan alınan kararlar

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
U1	50	60	90	55	45	45	80	75	65	55	55	70	65
U2	70	60	100	80	60	70	100	100	90	60	80	80	60
U3	80	50	80	30	60	40	60	70	60	40	40	60	50
U4	60	40	50	20	20	20	50	40	40	10	30	50	10
U5	20	40	70	30	30	20	60	80	60	50	50	40	60
U6	60	40	50	30	70	60	50	70	50	30	30	40	50
U7	90	80	70	40	20	40	70	80	70	50	35	50	60
U8	90	90	20	15	10	20	70	90	40	70	45	30	68
U9	80	60	50	30	60	25	40	60	30	30	30	10	10
U10	80	80	20	20	10	10	60	70	40	40	20	10	40
U11	70	70	90	40	60	70	90	70	30	70	60	75	90
U12	90	40	60	40	40	20	60	60	10	40	40	20	40
U13	90	90	80	20	40	50	95	100	60	30	40	100	70
U14	65	60	50	5	10	10	20	60	10	15	15	10	10
U15	50	50	30	10	20	10	20	50	5	10	10	20	20
U16	40	40	25	10	10	10	45	40	70	10	5	20	40
U17	100	80	75	60	85	80	85	90	60	70	55	70	60
U18	70	50	50	20	40	20	60	70	20	5	5	20	40
U19	80	50	50	20	30	50	60	70	70	10	20	60	20

	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26
U1	70	80	70	60	65	70	65	75	45	55	65	80	75
U2	70	50	50	60	50	40	60	40	30	50	40	70	30
U3	80	60	80	90	60	50	60	50	30	40	30	70	40
U4	40	20	50	40	40	10	10	30	20	10	10	50	30
U5	70	50	70	70	70	60	70	70	60	60	50	60	70
U6	50	50	60	50	60	40	40	60	40	50	70	60	40
U7	65	70	45	60	40	30	40	45	25	35	20	20	25
U8	95	50	80	95	80	60	75	30	10	25	10	50	15
U9	20	10	15	30	20	10	10	20	15	10	30	40	10
U10	40	40	60	50	40	20	50	60	50	50	30	70	50
U11	80	30	70	70	70	50	70	60	55	40	40	60	50
U12	20	10	20	40	40	40	10	10	10	10	20	20	10
U13	70	20	80	95	50	40	80	90	30	40	60	80	50
U14	55	10	25	5	15	5	40	10	10	5	20	40	0
U15	50	10	20	60	10	5	5	5	5	5	20	50	10
U16	10	40	35	70	50	5	25	55	40	3	50	60	10
U17	70	65	80	60	70	60	75	60	55	60	65	70	70
U18	50	80	70	80	50	25	65	40	20	50	50	90	15
U19	30	50	50	60	30	20	60	50	30	30	10	50	10

Ek 4. Nedelsky yöntemi ile uzmanlardan alınan kararlar

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
U1	33	33	20	33	33	50	20	20	20	33	20	50	20
U2	50	50	20	50	50	50	20	20	20	20	20	50	20
U3	50	33	50	33	50	33	50	50	50	33	25	50	33
U4	33	33	33	20	33	33	20	20	20	20	20	20	20
U5	33	50	20	25	33	25	20	20	20	20	20	25	20
U6	50	33	50	25	50	50	33	50	33	33	33	33	33
U7	100	50	20	20	33	50	20	20	20	50	33	50	20
U8	100	100	20	20	25	25	50	100	25	50	25	33	33
U9	50	50	33	50	50	33	33	100	20	33	33	20	33
U10	50	50	33	20	50	20	33	20	20	20	20	50	20
U11	50	50	20	20	50	25	20	20	20	20	20	50	20
U12	50	33	50	20	33	20	33	50	20	20	20	20	20
U13	50	100	100	20	50	33	100	100	25	25	20	100	100
U14	100	50	50	25	25	25	50	33	33	25	25	33	20
U15	33	33	33	25	33	33	33	50	33	33	25	33	50
U16	33	50	20	20	33	20	100	20	20	20	20	50	20
U17	100	50	25	20	50	50	20	20	20	20	20	100	20
U18	50	50	25	20	33	20	33	50	25	20	20	33	33
U19	50	50	100	25	33	33	20	20	100	20	20	50	20

	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26
U1	33	33	20	20	25	25	33	20	20	50	20	20	20
U2	33	50	33	20	33	33	33	20	20	50	33	33	33
U3	50	50	50	100	33	33	33	33	25	33	25	33	25
U4	33	20	50	20	20	20	33	20	20	20	33	50	20
U5	33	33	20	20	20	20	20	20	20	20	20	25	20
U6	33	33	33	33	33	33	33	50	33	33	50	50	33
U7	100	20	20	100	100	20	100	20	100	20	50	20	20
U8	100	33	50	100	50	25	33	25	20	25	25	25	20
U9	33	25	50	50	33	33	50	50	50	33	50	100	33
U10	50	50	25	20	50	20	25	20	50	20	25	50	20
U11	33	20	20	25	25	100	20	20	20	33	33	33	20
U12	33	33	20	33	33	20	20	33	20	20	20	25	20
U13	100	25	50	100	33	25	50	100	25	25	50	100	25
U14	100	25	50	20	33	25	50	25	100	25	50	50	20
U15	100	33	50	100	33	25	25	33	33	33	50	50	25
U16	20	33	20	100	100	20	20	20	100	20	20	100	20
U17	100	100	20	100	100	20	100	20	100	20	100	100	20
U18	33	50	33	100	33	25	33	33	20	25	33	33	25
U19	33	25	20	20	20	20	50	33	33	20	20	100	20

Ek 5. Ebel yöntemi ile uzmanlardan alınan kararlar

	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13
U1	60	65	85	60	45	45	75	75	75	65	60	85	75
U2	70	70	85	80	55	70	85	85	70	45	70	80	45
U3	80	40	60	40	40	30	80	60	40	30	40	40	50
U4	50	50	50	30	10	10	50	40	35	30	10	40	40
U5	20	50	60	55	50	20	55	70	50	55	55	40	60
U6	70	60	50	60	70	65	60	70	70	60	50	50	60
U7	90	80	55	35	20	55	65	65	70	55	30	50	35
U8	80	80	30	20	20	30	50	80	40	60	30	50	60
U9	60	30	20	20	20	20	30	60	20	25	25	20	10
U10	60	60	30	50	40	10	40	50	40	40	20	10	30
U11	70	40	70	70	50	70	70	70	40	40	40	50	70
U12	90	40	90	10	10	40	10	90	10	10	10	10	10
U13	70	70	70	30	50	50	80	100	50	25	30	100	50
U14	60	60	60	10	10	10	20	55	10	10	10	10	10
U15	50	30	25	10	25	10	30	50	5	10	10	25	25
U16	40	40	10	10	10	10	45	40	10	10	10	10	40
U17	65	65	60	60	65	60	65	65	55	60	55	70	55
U18	80	65	50	15	60	20	50	75	20	15	15	20	65
U19	80	50	50	20	20	50	65	65	65	20	20	65	20

	M14	M15	M16	M17	M18	M19	M20	M21	M22	M23	M24	M25	M26
U1	75	80	60	60	65	65	65	85	50	50	60	85	85
U2	70	45	55	85	85	55	55	45	45	70	45	55	45
U3	80	50	60	80	50	30	30	30	40	30	30	60	40
U4	30	40	40	40	40	10	10	40	10	40	10	40	10
U5	70	50	70	70	70	55	70	70	55	55	50	60	55
U6	50	60	40	60	70	50	50	50	50	50	70	60	40
U7	55	55	65	65	35	30	55	55	35	30	30	35	35
U8	80	50	60	80	80	60	70	40	50	30	30	40	20
U9	30	10	20	20	20	25	20	20	30	20	30	20	30
U10	50	50	60	40	30	10	50	50	50	50	40	60	40
U11	70	40	70	70	70	50	50	70	50	50	40	70	40
U12	10	40	10	40	40	10	10	10	10	10	10	10	10
U13	50	25	80	80	50	50	80	80	25	50	50	70	50
U14	55	10	35	10	10	10	20	10	10	10	20	25	10
U15	55	10	25	55	10	10	10	10	10	5	30	50	10
U16	10	40	40	70	40	10	10	40	40	10	40	60	10
U17	60	55	70	60	70	60	70	60	60	60	60	70	65
U18	50	80	75	75	65	20	70	60	15	50	60	80	10
U19	20	50	50	65	20	20	50	40	40	20	20	50	20