

Angoff, Nedelsky ve Ebel Standart Belirleme Yöntemleri ile Belirlenen Kesme Puanlarının Karşılaştırılması*

A Comparison of Cutting Points Determined by Angoff, Nedelsky and Ebel Standard Setting Methods

Okan DEMİR**, İ. Alper KÖSE***

Özet: Bu araştırma ile test merkezli standart belirleme yöntemlerinden 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 araştırmacının öğrenci verileri olarak kullanılmıştır. Uzman verileri ise araştırmacı tarafından geliştirilen “Uzman Görüşleri Formu” ile toplanmıştır. Başarılı kabul edilen öğrenci yüzdeleri arasında anlamlı fark olup olmadığı bağımlı iki yüzde arasındaki farkın testi ile sınanmıştır. Yöntemler arasındaki uyum Cohen’s Kappa İstatistiği ile, MGP’ler arasındaki ilişki Pearson Momentler Çarpımı Korelasyon Katsayısı ile, MGP’lerin ortalamaları arasındaki fark ise Bağımlı Gruplar t Testi ile incelenmiştir. Araştırma sonucunda en yüksek kesme puanının Angoff yöntemi, en düşük kesme puanının Nedelsky yöntemi tarafından belirlendiği, başarılı kabul edilen öğrenci yüzdelerinin anlamlı şekilde farklılaştığı görülmüştür. Yöntemler tarafından belirlenen kesme puanlarına göre öğrenciler başarılı-başarısız şeklinde kategorize edildiğinde, 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, MGP’ler bakımından düşük düzeyde bir ilişki olduğu görülürken, MGP’lere ait ortalamalar arasındaki fark anlamlı bulunmuştur. Angoff ve Ebel yöntemleri arasında yüksek düzeyde anlamlı bir ilişki olduğu belirlenmiş, MGP’lere ait ortalamalar arasındaki fark anlamlı bulunmamıştır. Nedelsky ve Ebel yöntemleri arasında düşük düzeyde bir ilişki olduğu gözlenmiş, MGP’lerin ortalamaları arasındaki fark anlamlı bulunmamıştır.

Anahtar Kelimeler: Angoff, ebel, nedelsky, kesme puanı.

Abstract: In this study, it was aimed to compare the cutting points determined by Angoff, Nedelsky and Ebel methods. Data were collected from two different groups comprising of students and experts. The mathematics achievement test, comprising of 26 questions which was developed by the researcher with the aim of evaluating the objectives of 9th grade mathematics course, was applied to 245 students and obtained achievement scores were used as student data of the study. Expert data of the study were collected by the “Expert Opinion Form” developed by the researcher. The differences between the percentages of students, considered as successful, have been tested with the test of difference between two percentages. The compliance between three methods has been examined with Cohen’s Kappa statistics. The relationship between the MGP 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 while the lowest cut-off score is determined by the Nedelsky and also the percentages of students differ statistically significant. When the students are classified as successful and

* Bu çalışma, Okan DEMİR tarafından Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü’nde Yrd. Doç. Dr. İ. Alper KÖSE danışmanlığında yapılan yüksek lisans tez çalışmasından özetlenmiştir.

** Ölçme ve Değerlendirme Uzmanı, İzzet Baysal Tek. ve End. Mes. Lis., Bolu, e-posta: okkandemir@gmail.com

*** Yrd. Doç. Dr. İ. Alper KÖSE, Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, Bolu, e-posta: i.alper.kose@gmail.com

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

unsuccessful it is seen that there is a high positive compliance between Angoff and Ebel, and Ebel and Nedelsky while there is a positive compliance between Angoff and Nedelsky. In terms of the MGP, there is compliance between Angoff and Ebel, but there is no compliance between Angoff and Nedelsky.

Key Words: Angoff, nedelsky, ebel, cut-off score.

GİRİŞ

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 ölçütü 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 üniteye 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 kararlar verildiği, bazen de geçerli olmayan ölçütler seçildiği veya ölçütün yanlış kullanıldığı, hatta bazen hiç ölçüt seçilmeden kararlar verildiği görülmektedir. Bu tür yanlışlar isabetli olmayan kararları beraberinde getirdiğinden dolayı, 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). Süreç sonunda elde edilen kesme puanı ölçme sonuçlarına bağlı olarak, bireyleri iki veya daha fazla performans düzeyine yerleştirmek amacıyla kullanılır (Taşdelen, 2009). Bazen de bu kararlar sertifika kararları, yerleştirme kararları, teşvik kararları ve lisanslama kararları şeklinde olabilir. 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 sınav sonuçlarına 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 Yöntemi olarak bilinmektedir. Bu araştırma kapsamında incelenen test merkezli yöntemlerden Angoff, Nedelsky ve Ebel yöntemlerinden aşağıda kısaca bahsedilmiştir.

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 1’de verilmiştir.

Tablo 1. 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 1’de 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, NEDELSKY VE EBEL STANDART BELİRLEME YÖNTEMLERİ İLE BELİRLENEN KESME PUANLARININ KARŞILAŞTIRILMASI

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'de Nedelsky yöntemi ile kesme puanı belirlemeye yönelik bir örnek verilmiştir.

Tablo 2. *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	$\frac{1}{2}=.50$	$\frac{1}{4}=.25$	$\frac{1}{2}=.50$	$\frac{1}{3}=.33$	$\frac{1}{1}=1$	$\frac{1}{5}=.20$	$\frac{1}{3}=.33$	$\frac{1}{1}=1$	$\frac{1}{5}=.20$	$\frac{1}{1}=1$	5.31

Tablo 2'de 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 seçenektan 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.

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 yanıtlanma 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 3'te gösterilmiştir.

Tablo 3. Ebel Yöntemi ile “Zorluk” ve “Uygunluk” Boyutlarının Tespit Edilmesini Gösteren Bir Örnek

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

Tablo 3'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 4'te Tablo 3'ten yararlanarak kesme puanı hesaplama süreci gösterilmektedir.

Tablo 4. 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
Uzmana ait minimum geçme puanı			12.05

Tablo 4'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

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

minimum geçme puanı hesaplanmıştır. Tüm uzmanların minimum geçme puanlarının ortalaması ise testin nihai kesme puanını vermektedir.

Bu üç yöntemden 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. Araştırma kapsamında 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?

Araştırmanın Amacı ve Önemi

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.

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; Gündeğer ve Doğan, 2014; Ömür ve Selvi, 2010; 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 (Ömür ve Selvi, 2010; Poggio ve diğerleri, 1982). 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.

YÖNTEM

Bu araştırma ile farklı üç standart belirleme yöntemine göre elde edilen kesme puanları karşılaştırılmış, MGP'ler arasındaki ilişkiler incelenmiştir. Bu açıdan çalışma, ilişkisel bir araştırmadır. Araştırma ile yöntemlerin belirlediği kesme puanları, durum saptamaya yönelik incelendiğinden, çalışma bu yönüyle betimsel bir araştırma niteliği taşımaktadır.

Çalışma Grubu

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 amaçlı örnekleme yöntemine göre yapılmıştır.

Karar verme aşamasında görüşlerine başvurulacak uzman grubun sayısı standart belirleme yöntemlerinde en çok tartışılan konulardan biridir. Her ne kadar uzman sayısına bir sınırlama getirilmese de Jeager'a göre (1989) 10 ile 15 arasında olan bir uzman sayısı yeterli görülmektedir. 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 amaçlı örnekleme yöntemiyle, gönüllülük esasına göre yapılmıştır. Uzman grubu alanında deneyimli matematik öğretmenlerinden oluşmaktadır.

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 alınan öğrenci puanlarıdır. Öğrencilerden veri toplama aşamasında, 9. Sınıflar Matematik Dersi Öğretim Programı 1. dönem müfredatında bulunan Kümeler ile Denklem ve Eşitsizlikler öğrenme alanlarını kapsayan 34 maddelik bir başarı testi hazırlanmıştır. Hazırlanan bu denemelik test 490 öğrenciye bir ders saati süresince uygulanmış, uygulama sonucu elde edilen veriler Microsoft Excel (2007)'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, madde seçimi yapılarak 26 maddelik başarı testi oluşturulmuştur. Denemelik testin iç tutarlık güvenirlik katsayısı (KR-20) .95 olarak kestirilmiştir. Hazırlanan başarı testi çiftler olarak adlandırılan ikinci gruba uygulanmış, bu uygulama ile elde edilen veriler öğrenci verileri olarak değerlendirilmiştir. Nihai uygulama sonucu oluşan KR-20 değeri .93 olarak bulunmuştur.

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 hazırlanmıştır. Araştırmacı tarafından hazırlanan "Uzman Görüşleri Formu" yardımıyla 19 matematik öğretmeninden görüş alınmış ve her uzmanın her bir maddeye verdikleri kararlar doğrultusunda her bir yöntem için bir kesme puanı belirlemiştir. 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.

Verilerin Analizi

Üç farklı standart belirleme yöntemine göre belirlenen kesme puanlarının üzerinde puan alarak, başarılı kabul edilen öğrenci yüzdeleri arasında anlamlı bir farklılığın olup olmadığını araştırılmasında, bağımlı iki yüzde arasındaki farkın testi ve z istatistiği kullanılmıştır. Yöntemler arasında öğrencilerin başarılı-başarısız olarak sınıflandırılma durumları bakımından uyumun araştırılmasında Cohen'in Kappa Katsayısı kullanılmıştır. Yöntemler arasında uzmanlardan elde edilen minimum geçme puanları (MGP) bakımından uyumun araştırılmasında Pearson Momentler Çarpımı Korelasyon Katsayısı ve Bağımlı Gruplar t Testi kullanılmıştır. 19 uzmandan elde edilen MGP'ler arasındaki ilişki Pearson Momentler Çarpımı Korelasyon

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

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

BULGULAR

Üç farklı standart belirleme yöntemine göre belirlenen kesme puanlarının üzerinde puan alarak, başarılı kabul edilen öğrenci yüzdeleri arasında anlamlı bir farklılığın olup olmadığını araştırılmasında, 26 maddeden oluşan matematik başarı testinin her bir maddesi için 19 uzman görüşlerine başvurulmuştur. Her uzmanın her bir yöntem için verdiği kararlar toplanıp aritmetik ortalaması alınarak her uzmanın yöntemlere göre MGP'si hesaplanmıştır. Farklı yöntemlerle belirlenen kesme puanları ve bu kesme puanlarına göre başarılı kabul edilen öğrenci sayısı ve yüzdeleri aşağıda verilmiştir.

Tablo 5. 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 5 incelendiğinde, Matematik başarı testi için en yüksek kesme puanının 46,39 değeriyle Angoff yöntemi tarafından elde edildiği, Angoff yöntemini 44,89 değeriyle Ebel yönteminin takip ettiği, en düşük kesme puanının ise Nedelsky yöntemi tarafından 37,71 olarak belirlendiği görülmüştür. Başarılı kabul edilen öğrenci yüzdesinin ise en çok Nedelsky yönteminde (%57), en az ise Angoff yönteminde (%43) olduğu görülmüştür. Her bir yöntemle göre başarılı kabul edilen öğrenci yüzdeleri arasındaki farklılıkların anlamlılığı bağımlı iki yüzde arasındaki farkın testi ile sınanmış olup sonuçları Tablo 6'da verilmiştir.

Tablo 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 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.

Öğrencilerin başarılı-başarısız olarak sınıflandırılmaları bakımından farklı standart belirleme yöntemleri arasındaki uyumu belirlemek amacıyla Cohen'in Kappa testi kullanılmıştır. 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 dönüştürülmüştür. Kappa testi sonuçları aşağıda verilmektedir.

Tablo 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 aldığı makul bir uyumdan, 0,75'ten büyük bir değer aldığı ise mükemmel bir uyumdan söz edilebilir (Şencan, 2005). Tablo 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.

Yöntemler arasında uzmanlardan elde edilen minimum geçme puanları (MGP) bakımından uyumun araştırılması 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 8'de verilmiştir.

Tablo 8. MGP'lere Ait Betimsel İstatistikler 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 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 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 bakımından 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 9'da verilmiştir.

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

Angoff-Ebel	0,963	0,00*	19
Nedelsky-Ebel	0,088	0,722	19

*p<.001

Tablo 9 incelendiğinde, Angoff ve Nedelsky yöntemleri ile Nedelsky ve Ebel yöntemleri arasında uzmanlardan elde edilen MGP'ler bakımından düşük düzeyde bir ilişki olduğu görülürken, Angoff ve Ebel yöntemleri arasında ise yüksek düzeyde pozitif bir ilişki olduğu görülmektedir. 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 10'da verilmiştir.

Tablo 10. *MGP'lere Ait Ortalamalar Arası Farkın Anlamlılığı İçin Yapılan T Testi Sonuçları*

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

*p<.01

Tablo 10 incelendiğinde Angoff ve Nedelsky yöntemlerine ait MGP'lerin ortalamaları arasındaki farkın anlamlı olduğu görülmektedir [$t_{(18)}=0,033$, $p<.05$]. Angoff ve Ebel yöntemlerine ait MGP'lerin ortalamaları arasındaki fark [$t_{(18)}=1,568$, $p>.05$] ile, Nedelsky ve Ebel yöntemlerine ait MGP'lerin ortalamaları arasındaki fark [$t_{(18)}=-1,828$, $p>.05$] ise anlamlı değildir.

TARTIŞMA, SONUÇ VE ÖNERİLER

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. Elde edilen bulgulara göre, 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. 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. 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.

Bu araştırma ile, her bir yöntemle 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. 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 kolay-zor şeklinde kodlayarak öğrencilerin maddeleri yapabilirlikleri hakkında karar verirken, Nedelsky yönteminde uzmanlar soruların kolay ve zorluğundan ziyade seçeneklerin elenebilirliğine odaklanmaktadır. 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 doğurmaktadır.

Uzman kararlarından elde edilen MGP'ler bakımından, Angoff ve Nedelsky yöntemleri arasında ilişki olmadığı, MGP'lerin ortalamalarının anlamlı şekilde farklı olduğu 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.

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 benzer bilimsel çalışmaların bulguları dikkate alınarak yeniden belirlenmesi yönünde fikir beyan edilebilir.

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 için de yapılabilir.

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

KAYNAKLAR

- Appleton, L. P. (1984). *A comparison of the nedelsky, ebel, and contrasting groups methods of cut score determination*. Yayınlanmamış Doktora Tezi. The Catholic University of America, ABD.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- 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., & Lockwood, R. E. (1980). A comparison of the nedelsky and angoff cutting score procedures using generalizability theory. *Applied Psychological Measurement*, 4(2), 219 – 240.
- 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.
- Cizek, G. J. (2001). *Setting performance standards: Concepts, methods and perspectives*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Cizek, G. J. & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. California: Sage Publication.
- Crocker, L. & 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*. Yayınlanmamış Doktora Tezi. Hacettepe Üniversitesi, Ankara.
- Ertürk, S. (1994). *Eğitimde program geliştirme* (8. Baskı). Ankara: Meteksan Yayınevi.
- Gündeğer, C. (2012). *Angoff, yes/no ve ebel standart belirleme yöntemlerinin karşılaştırılması*. Yayınlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Ankara.
- Gündeğer, C. ve Doğan, N. (2014). Angoff, yes/no ve ebel standart belirleme yöntemlerinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(1), 53-60.
- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. G. J. Cizek (Edt). *Setting performance standards: Concepts, methods and perspectives* (ss. 89-116). Mahwah, NJ: Lawrence Erlbaum Associates.
- Irwin, P. (2007). *An alternative examinee-centered standard setting strategy*. Yayınlanmamış Doktora Tezi. University of Nebraska, ABD.
- Jeager, R. M. (1989). Certification of student competence. R. L. Linn (Edt). *Educational measurement* (ss. 485-514). 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.
- Karasar, N. (1982). *Bilimsel araştırma yöntemi*. Ankara: Matbaş Matbaacılık ve Ambalaj Sanayi.
- Kubiszyn, T., & Borich G. (2007). *Educational testing and measurement classroom application and practice*. USA: Wiley Jossey-Bass Education.
- Livingston, S. A., & Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests*. New Jersey: Educational Testing Service. 10 Ocak 2014 tarihinde http://www.ets.org/Media/Research/pdf/passing_scores.pdf adresinden alınmıştır.
- Mills, C. N., & Melican, G. J. (1988). Estimating and adjusting cutoff scores: Features of selected methods. *Applied Measurement in Education*, 1(3), 261 – 275.
- Ömür, S. ve Selvi, H. (2010). Angoff, ebel ve nedelsky yöntemleriyle belirlenen kesme puanlarının sınıflama tutarlıklarının karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 109-113.
- Poggio, J. P., Glasnapp, D. R. ve Eros, P. I. (1982). An Evaluation of Contrasting Groups Methods for Setting Standarts. Paper Presented at the Annual Meeting of the American Educational Research Association. New York: USA.
- Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenilirlik ve geçerlilik*. Ankara: Seçkin Yayıncılık.
- Tanrıverdi, S. (2006). *Standart belirleme yöntemlerinin geçme puanları üzerine etkisi*. Yayımlanmamış Yüksek Lisans Tezi, Ankara.
- Taşdelen, G. (2009). *Nedelsky ve angoff standart belirleme yöntemlerinin genellenebilirlik kuramı ile karşılaştırılmasına ilişkin bir araştırma*. Yayımlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Ankara.

- Taşdemir, F. (2013). *Angoff (1-0), nedelsky ve sınır değerleri saptama yöntemleri ile bir testin sınıflama doğruluklarının incelenmesi*. Yayımlanmamış Doktora Tezi. Ankara Üniversitesi, Ankara.
- 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*. Yayımlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Ankara.

EXTENDED ABSTRACT

Introduction

In determining the student achievement, elections of student or being accepted for a job, various tests are utilized. In accordance with the aim of test, these tests are prepared to find out the individual characteristics desired. In many cases, to interpret the result of any test we primarily need to obtain a cut off score. For instance, in a subject that allocated to the units, for passing the from one unit to another students are expected to reach a certain level of knowledge. This knowledge level is determined by achieve or cut off score. If individuals receive a score above the cut off point may pass the next unit or If individuals receive a score below the cut off point it is expected for the individual to take the same test. Sometimes it is seen that practitioners make a decision on the basis of unreliable measurement results. It is observed that sometimes invalid criterias are selected or criterias are utilized incorrectly also sometimes it seen that decisions are made without selecting a criteria. On account of this the criteria or standart setting process is very important. Standard setting is a process to determine the performance point that seperate the performance levels and be termed as a cutting or passing score. In the literature there are many standart setting methods and determining the which method will be used is important. When we scan the international literature that is accessible, there are many studies that compare different methods of standard setting. However, there are no enough studies that compare standard setting methods in the national literature. It is seen that number of studies that deal with the Ebel method are quite a few than another studies. There are very few studies that deal with Angoff, Nedelsky and Ebel Methods at the same time in the literature. In this context, with this study, it was aimed to compare the cutting points determined by Angoff, Nedelsky and Ebel methods, one of test-based standard setting methods.

Method

With this study, cut-off scores determined by three different standard setting methods were compared and relationship between MGP obtained were examined. In this regard this study is a relational research. Research models that aim to determine the presence or extent of covariance change between two or more variables, are defined as a kind of relational research models (Karasar, 1982, p. 85). The study is a descriptive study because of determining the case of cut-scores obtained by 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 were 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, was applied to 245 students and obtained achievement scores were used as student data of the study. Expert data of the study were 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.

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

Result and Discussion

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.