

Angoff, Yes/No ve Ebel Standart Belirleme Yöntemlerinin Karşılaştırılması*

A Comparison of Angoff, Yes/No and Ebel Standard Setting Methods

Ceylan GÜNDEĞER**

Nuri DOĞAN***

Öz

Bu araştırmada Angoff, Yes/No ve Ebel standart belirleme yöntemleri ile elde edilen uzman kararları ve kesme puanları karşılaştırılmıştır. Araştırmanın verilerini, 20 maddelik bir başarı testinden alınan 489 öğrenci puanı ve 17 uzmanın bu başarı testindeki maddeler için üç farklı standart belirleme yöntemine göre vermiş oldukları kararlar oluşturmaktadır. Verilerin analizi uzmanlar arası uyum Angoff ve Ebel yöntemleri için Kendall'in uyum katsayısı (W) ve Yes/No yöntemi için Cohen'in Kappa istatistiği ile; üç farklı standart belirleme yöntemine göre başarılı sayılan öğrenci yüzdeleri arasındaki farklılıklar bağımlı iki oran arasındaki farkın testi ile; üç yöntemden elde edilen kesme puanlarına göre öğrenci puanları başarılı-başarısız olmak üzere yapay ikili hâle getirilip aralarındaki uyuma Cohen'in Kappa istatistiği ile; üç yöntemde ait kesme puanları arasındaki uyum ise Pearson Momentler Çarpımı Korelasyon Katsayısı ve bağımlı gruplar t testi ile test edilmiştir. Bu araştırma sonunda yöntemlerde uzmanlar arasında makul bir uyumun yakalanabildiği; üç farklı standart belirleme yöntemine göre başarılı sayılan öğrenci yüzdeleri arasında 0,01 hata düzeyinde manidar farklılık olduğu ve öğrencilerin başarılı-başarısız sayılma durumları bakımından uyum olduğu görülmüştür. Kesme puanları arasında Angoff-Yes/No yöntemleri arasında yüksek düzeyde, Angoff-Ebel yöntemleri arasında orta düzeyde bir ilişki olduğu ve Yes/No-Ebel yöntemleri arasında anlamlı bir ilişki olmadığı görülmüştür.

Anahtar Kelimeler: Standart belirleme yöntemleri, angoff yöntemi, yes/no yöntemi, ebel yöntemi, kesme puanı

Abstract

In this study, experts' decisions and cut-off scores obtained from Angoff, Yes/No and Ebel standard setting methods have been compared. The research data consist of 489 student scores obtained from a achievement test of 20 items and 17 experts' decisions, made according to three different standard setting methods for this achievement test. In the analysis of data, in order to set the compliance between experts, Kendall's coefficient of concordance (W) for Angoff and Ebel methods, and Cohen's Kappa statistics for Yes/No method have been calculated. Differences between the percentages of students, considered as successful according to three different standard setting methods, have been tested with the test of the difference between two correlated proportions. The students' scores have been tailored to binary as successful-unsuccessful according to three methods; and the compliance between them has been examined with Cohen's Kappa statistics. The compliance between the cut-off scores of three methods has been evaluated with Pearson Product-Moment Correlation Coefficient and the paired samples t test. At the end of this research, it is seen that the compliance of experts in the methods is at a reasonable level; there is significant difference on 0,01 error level among percentages of students considered as successful according to these three different standard setting methods and there is a compliance between students' scores that have been tailored to binary as successful-unsuccessful according to three methods. It is concluded that there is a high relationship between Angoff and Yes/No methods; there is moderate relationship between Angoff and Ebel methods; there is no significant relationship between Yes/No and Ebel methods.

Key Words: Standard setting methods, angoff method, yes/no method, ebel method, cut-off score

* *Yazarın yüksek lisans tezinin bir kısmını içeren bu çalışma, III. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sözlü bildiri olarak sunulmuştur.

** Arş. Gör., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, cgundeger@gmail.com

*** Doç. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, nuridogan2004@gmail.com

GİRİŞ

Tüm dünyada olduğu gibi ülkemizde de öğrenci başarısının belirlenmesinde, işe alımlarda veya öğrenci seçimlerinde çeşitli testlerden yararlanılmaktadır. Bu testler, amaca uygun olarak, bireyin istenilen özelliklere ne kadar sahip olduğunu ortaya çıkaracak şekilde hazırlanmaktadır. Bireyin istenilen özelliklere ne derece sahip olduğu testlerle belirlendiği gibi, bireyler hakkındaki kararlar da yine test sonuçları dikkate alınarak verilmektedir. Bireyler hakkında karar verilirken bireylere ait başarı, çeşitli ölçütlerle kıyaslanarak değerlendirilmektedir. Değerlendirmelerde kullanılan ölçütler bağıl veya mutlak olabilmekte ve standart belirleme yöntemleri ile belirlenebilmektedir.

Standart belirleme süreci, kişileri performans düzeylerine göre ayıran performans standartlarını, test puanları cetvelinde rakamsal hâle getirme süreci olarak düşünülebilir (Hambleton, 2001). Son yıllarda alanyazında otuzdan fazla standart belirleme yöntemi betimlenmesine karşın, bu yöntemleri genel olarak (i) Bütünsel Etkiye Dayanan Yöntemler, (ii) Test Maddelerine Dayanan (*Test Merkezli*) Yöntemler ve (iii) Testi Alanların Performansına Dayanan (*Performans Merkezli*) Yöntemler olmak üzere üç ana başlık altında toplamak mümkündür (Crocker ve Algina, 1986).

Bu araştırma kapsamında test merkezli yöntemlerden Angoff, Yes/No ve Ebel yöntemleri karşılaştırılmıştır. Test merkezli yöntemlerde standart belirleme sürecinde, “*Ne kadar yeterlidir ?*” sorusuna cevap verecek olan uzman grubunun kararlarından faydalanılmaktadır. Her uzman bireysel olarak, belirlenen yönteme uygun şekilde testle ilgili kararlarını verdikten sonra her bir uzmana ait minimum geçme puanı (*minimum passing level*) (MGP) hesaplanır. Bu puanların ortalaması alınarak testin nihai geçme puanı belirlenir (Livingston ve Zieky, 1982).

Angoff Yöntemi

Angoff yönteminde uzmanlar, maddeyi bir bütün olarak düşünür ve geçme-kalma sınırında minimum yeterliğe sahip öğrencilerin (sınır grubun) maddeyi doğru cevaplama olasılığını belirler. Sorular kolaylaştıkça bu olasılık artacaktır. Eğer herhangi bir çoktan seçmeli test maddesi için olasılık sıfır olarak düşünülüyse, o test maddesine ait şans başarısı olan “*1/seçenek sayısı*” olasılık olarak kabul edilmelidir. Maddelere verilen bu yüzdeler toplanarak uzmanlara ait MGP’ler hesaplanır (Livingston ve Zieky, 1982).

Yes/No Yöntemi

Yes/No yönteminde uzmanlardan testteki maddeleri inceleyerek, sınır gruptaki bir öğrencinin maddeyi doğru cevaplayıp cevaplamayacağına karar vermeleri istenir. Uzman tarafından, sınır grup öğrencisinin maddeyi doğru cevaplayacağı düşünülüyorsa maddeye 1, yanlış cevaplayacağı düşünülüyorsa maddeye 0 (sıfır) puan verilerek, her bir uzmana ait MGP hesaplanır (Zieky, 2001).

Ebel Yöntemi

Ebel yönteminde uzmanların diğer yöntemlerden farklı olarak her bir madde için iki farklı karar vermeleri gerekmektedir. Bu kararlardan biri her madde için zorluk ve diğeri ise uygunluk düzeyinin tahmin edilmesidir. Bu iki aşama kendi içinde “*Zor, Orta, Kolay*” ve “*Gerekli, Önemli, Kabul Edilebilir, Tartışılabilir*” olarak derecelendirilmiştir (Ebel, 1972). Ebel yönteminde ilk olarak, uzmanlardan maddeleri *sadece* testin yapısını dikkate alarak 3x4’lük sınıflama tablosuna uygun şekilde yerleştirmeleri istenir. Sonrasında ise sınır grup öğrencilerini dikkate alarak bu sınıflamalara ait doğru cevaplanma yüzdelerinin belirlenmesi gerekmektedir. Bu iki aşama tamamlandıktan sonra, her hücrede bulunan madde sayısı ile o

hücreye ait yüzde çarpılarak her hücrenin puanı belirlenir ve belirlenen bu puanlar toplanarak uzmanlara ait MGP'ler hesaplanır (Livingston ve Zieky, 1982).

Alanyazında bütün testler için genel geçer bir “*en iyi yöntem*” tanımlanmamıştır. Test merkezli yöntemlerin hepsi uzmanların karar verme sürecini gerektirir ve bu kararlar öznelidir (Downing, Tekian ve Yudkowsky, 2006). Standart belirleme yöntemleri ile elde edilen ölçütlere göre öğrenci başarıları, öğrencilerin geçti-kaldı ve/veya başarılı-başarısız olma durumları belirlenmektedir. Öğrenci hakkında karar verme aşamasında bir ölçüt olan standartlar, aynı test için farklı standart belirleme yöntemlerine göre farklılık gösterebilmektedir.

Araştırmanın Amacı

Bu araştırmanın amacı Angoff, Yes/No ve Ebel standart belirleme yöntemleri arasında nasıl bir ilişki olduğunun belirlenmesidir. Bu amaçla öncelikle yöntemlerde uzman kararları arasındaki uyumun nasıl olduğu; kesme puanlarına göre başarılı kabul edilen öğrenci yüzdeleri arasında anlamlı bir farklılık olup olmadığı; kesme puanlarına göre öğrencilerin başarılı-başarısız olarak sınıflandırılma durumları bakımından uyumun nasıl olduğu belirlenmiş ve MGP'ler arasındaki uyuma bakılmıştır.

YÖNTEM

Bu araştırma üç farklı standart belirleme yöntemini karşılaştırmak amacıyla yapılmış ilişkisel türden bir araştırma olup aynı zamanda her bir yönteme ilişkin betimleyici istatistikleri elde etme, iki ya da daha fazla değişken arasındaki ilişkinin varlığını ve derecesini ortaya koyma yönleriyle de betimsel bir çalışmadır.

Çalışma Grubu

Araştırmada 489 ilköğretim 4. sınıf öğrencisi yer almış olup araştırmanın uzman grubunu ise alanında deneyimli 17 sınıf öğretmeni oluşturmuştur.

Veri Toplama Araçları

Öğrenciler ve uzmanlardan aynı test aracılığıyla veri toplanmıştır. Araştırmada kullanılan öğrenci verileri, ilköğretim 4. sınıf matematik dersi “Bölme ve Kesirler” alt öğrenme alanlarına yönelik araştırmacı tarafından hazırlanmış 20 maddelik bir başarı testinden alınan puanlardır. Araştırmada kullanılan diğer verileri oluşturan, uzmanların hazırlanan bu başarı testindeki maddeler için üç farklı standart belirleme yöntemine göre vermiş oldukları kararlardır.

Başarı testinin geliştirilmesi aşamasında, “Bölme ve Kesirler” alt öğrenme alanına ait 11 kazanıma yönelik 40 maddelik bir test oluşturulmuş ve ön uygulamada 272 ilköğretim 5. sınıf öğrencisine uygulanmış olup nihai teste madde seçiminde madde ayırt edicilik gücü ve madde güçlük indeksleri dikkate alınmıştır. Seçilen 20 maddeden oluşan nihai testin güvenilirliği için KR-20 iç tutarlılık güvenilirlik katsayısı 0,84 olarak kestirilmiştir. Murphy ve Davidshofer’a göre (1991) sınıf içi başarı testlerinde iç tutarlılık katsayısının en az 0,75 olması gerekmektedir. Bu bakımdan kestirilen değerinin yeterli olduğu söylenebilir. Testin kapsam geçerliği için 4 alan uzmanına ait görüşlerin uyumu Kappa istatistiği ile 0,82 olarak hesaplanmıştır. Buna göre testin kapsam geçerliğinin yüksek olduğu söylenebilir.

İşlem

Oluşturulan 20 maddelik başarı testi 489 ilköğretim 4. sınıf öğrencisine uygulanmış olup KR-20 iç tutarlılık güvenilirlik katsayısı 0,82 olarak hesaplanmıştır. Nihai testin geçerlik çalışması

için asıl uygulamanın yapıldığı gruptan ulaşılabilen 180 öğrencinin birinci dönem karne notları ile uygulama sonucunda aldıkları puanlar arasındaki korelasyon Spearman Brown Sıra Farkları Korelasyon Katsayısı ile hesaplanmış olup bu değer 0,62 ($p < 0,01$) olduğu görülmüştür. Bu durum uygulama sonuçları ile karne notları arasında 0,01 hata ile orta düzeyde anlamlı bir ilişki olduğunu göstermektedir.

Nihai testte yer alan 20 madde için 17 uzmandan üç standart belirleme yöntemine göre kararlar alınmıştır. Buna göre Angoff yöntemi için uzmanlardan, minimum yeterlik düzeyindeki 100 öğrenciden kaçının o test maddesini doğru cevaplayabileceğini düşünmeleri; Yes/No yöntemi için minimum yeterlik düzeyindeki bir öğrencinin soruları doğru cevaplayıp cevaplayamama durumlarından uygun olanını işaretlemeleri; Ebel yönteminde ise testteki tüm maddeleri bahsedilen 3×4 'lük tabloya yerleştirip hücrelerdeki maddeler için ortak bir doğru cevaplanma yüzdesi belirlemeleri istenmiştir.

Verilerin Analizi

Uzman kararlarının uyumuna Angoff ve Ebel yöntemleri için Kendall'in uyum katsayısı (W) ile, Yes/No yöntemi için kategorik verilerin uyumunu ölçen Cohen'in Kappa istatistiği ile; yöntemlere göre başarılı sayılan öğrenci yüzdeleri arasındaki fark bağımlı iki oran arasındaki farkın testi ile; yöntemlerin ikili uyumlarına Cohen'in Kappa istatistiği ile; her bir yöntem için uzman kararlarından elde edilen 17 adet MGP arasındaki ilişkiye, MGP'lerin yapısı dikkate alınarak Pearson Momentler Çarpımı Korelasyon Katsayısı ve bağımlı gruplar t testi ile bakılmıştır.

BULGULAR

Yöntemlere ait uzman kararları arasındaki uyumun belirlenmesi için amacıyla Angoff yöntemi için W katsayısı 0,38 ve kararların maddeler bazında ele alınmasıyla Ebel yöntemi için W değerinin 0,32 olduğu görülmüştür. Bu durumda uzmanlar arasında iki yöntemde de yüksek olmasa da kabul edilebilir bir uyum olduğu söylenebilir. Yes/No yöntemi için ise Cohen'in Kappa istatistiği 0,60 olarak hesaplanmış olup bu değer de orta derecenin üstünde makul bir uyum olduğu yorumu yapılabilir.

Standart belirleme yöntemleri ile belirlenen kesme puanlarına göre başarılı sayılan öğrenci yüzdeleri arasında anlamlı bir farklılık olup olmadığının belirlenmesi için başarılı sayılan öğrenci yüzdeleri ikili olarak bağımlı iki oran arasındaki farkın testi ile sınanmıştır. Tablo 1'e göre başarılı sayılan öğrenci yüzdeleri arasında 0,01 hata düzeyinde manidar farklılık olduğu görülmüştür.

Tablo 1. Kesme Puanları, Başarılı Sayılan Öğrenci Sayıları-Yüzdeleri ve z Değerleri

Karşılaştırılan Yöntemler	Kesme Puanı	Başarılı Sayılan Öğrenci Sayısı	Başarılı Sayılan Öğrenci Oranı	z
Angoff Yöntemi	51,47	327	0,67	5,85**
Yes/No Yöntemi	57,94	293	0,60	
Ebel Yöntemi	47,12	371	0,76	6,63**
Angoff Yöntemi	51,47	327	0,67	
Ebel Yöntemi	47,12	371	0,76	8,85**
Yes/No Yöntemi	57,94	293	0,60	

* $p < 0,01$

Başarılı sayılan öğrenci yüzdelerindeki bu manidar farklılıkların nedeni üç yöntemde de verilen kararların farklı dayanakları olmasına bağlanabilir. Angoff yönteminde uzmanlardan 100 kişilik geçme-kalma sınırındaki öğrenciden kaçının maddeyi doğru cevaplayabileceğini düşünmeleri istenirken; Yes/No yönteminde uzmanlardan geçme-kalma sınırındaki sadece bir öğrenciyi dikkate almaları istenmektedir. Ebel yönteminde ise bu iki

yöntemden tamamen farklı olarak uzmanlardan maddeleri önce Zorluk ve Uygunluk düzeylerine göre 3x4' lük bir tabloya yerleştirmeleri, sonrasında da her hücrede bulunan soru/sorular için geçme-kalma sınırındaki 100 öğrenciden kaçının soruyu doğru cevaplayabileceği düşünmeleri istenmektedir. Hücrelerdeki madde sayısı ile bu yüzdeler çarpılarak MGP'ler hesaplanmakta ve kesme puanı da MGP'lerin ortalaması olmaktadır. Böylece yöntemlerin süreçlerindeki farklılık kesme puanlarına yansımaktadır. Kesme puanlarındaki bu farklılık da geçen öğrenci yüzdelerini etkilemektedir. Ayrıca Impara ve Plake (1997) çalışmalarında uzmanların Yes/No yönteminde daha kolay karar verdiği sonucuna ulaşmışlardır. Bahsi geçen çalışmada uzmanlar Angoff yönteminde karar vermede 100 kişilik bir grubu düşünmenin daha zor olduğunu belirtmişlerdir. Uzmanların yöntemlere ait süreçlerdeki bu tip algıları da kararları etkileyebilir. Dolayısıyla kesme puanları ve başarılı sayılan öğrenci sayıları-yüzdeleri de bu durumdan etkilenebilir.

Yöntemler arasında öğrencilerin başarılı-başarısız olarak sınıflandırılma durumları bakımından uyumun belirlenmesi amacıyla öğrenci puanları üç yönteme ait kesme puanlarına göre 1-0 puanlanmış ve yöntemler arasındaki ikili uyumlara Cohen'in Kappa istatistiği ile bakılmıştır. Kappa istatistiği ile beraber Pearson ve Spearman korelasyon katsayılarının yer aldığı Tablo 2'ye göre tüm yöntemler arasında öğrencilerin başarılı-başarısız olarak sınıflanma durumları bakımından pozitif yönde ve 0,01 manidarlık düzeyinde uyum olduğu görülmektedir.

Tablo 2. *Yöntemler Arasındaki Uyum İçin Hesaplanan Korelasyon Katsayıları*

Yöntemler	Kappa	Spearman	Pearson
Angoff – Yes/No	0,85**	0,86**	0,86**
Angoff – Ebel	0,78**	0,81**	0,81**
Yes/No – Ebel	0,64**	0,69**	0,69**

*p<0.01

Tablo 2 incelendiğinde Angoff ve Yes/No yöntemleri ile Angoff ve Ebel yöntemleri arasında öğrencilerin başarılı-başarısız olarak sınıflanma durumları bakımından mükemmel bir uyum görülürken; Yes/No ve Ebel yöntemleri arasında makul bir uyum söz konusudur. Angoff ve Yes/No yöntemlerinin arasındaki bu mükemmel uyumun nedeni iki yöntemde de uzmanların sadece sınır grup öğrencisine odaklanıyor olmasından ileri gelebilir. Uzmanlar bu iki yöntemde başka bir görevle yükümlü değildir. Ayrıca Yes/No yöntemi, Angoff yönteminden türetilmiştir. Angoff ve Ebel yöntemleri arasındaki uyum ise yöntemlerde uzmanlardan sınır gruba dair yüzde alınıyor olması nedeniyle yüksek olabilir. Yes/No ve Ebel yöntemleri arasındaki makul ancak diğerlerine kıyasla daha düşük bu uyumun nedeni iki yöntemdeki sürecin birbirinden oldukça farklı işliyor olması olabilir.

Yöntemler arasında uzmanlara ait MGP'ler bakımından uyumun belirlenmesi için öncelikle MGP'lere ait betimsel istatistikler hesaplanmış ve normallik testi yapılmış; MGP'lerin normal dağılım gösterdiği görülmüştür. Bu nedenle analizde parametrik yöntemler olan Pearson Momentler Çarpımı Korelasyon Katsayısı ve bağımlı gruplar t testi kullanılmıştır. Tablo 3'te yöntemlere ait MGP'ler arasındaki ilişkiler ve anlamlılık düzeyleri gösterilmiştir.

Tablo 3. *MGP'ler Arasındaki Pearson Korelasyon Katsayıları*

	Angoff Yöntemi	Yes/No Yöntemi	Ebel Yöntemi
Angoff Yöntemi	-	0,50*	0,83**
Yes/No Yöntemi	-	-	0,47

*p<0.05

**p<0.01

Tablo 3'e göre Angoff ve Ebel yöntemlerine ait MGP'ler arasında 0,01 hata düzeyinde yüksek bir ilişkinin varlığı söz konusudur. Bu değer yüksek olmasının nedeni iki yöntemde de sınır gruba ait doğru cevaplama yüzdelerinin kullanılıyor olması olabilir. Angoff yönteminde tüm maddeler için bu yüzdeler ile kararlar verilirken; Ebel yönteminde maddeler Zorluk ve Uygunluk düzeylerine göre sınıflandırıldıktan sonra düzeyler için yüzdeler verilmektedir. MGP'ler arasındaki diğer bir manidar ilişkinin 0,05 hata düzeyinde Angoff ve Yes/No yöntemleri arasında olduğu görülmektedir. Buna göre, Angoff yönteminin bir varyasyonu olan Yes/No yöntemi ile Angoff yöntemi arasında MGP'ler bakımından, Ebel ve Angoff yöntemleri arasındaki ilişkiden daha düşük, orta derecede bir ilişki olduğu yorumu yapılabilir. Yes/No ve Ebel yöntemlerine ait MGP'ler arasındaki 0,47 değerindeki ilişki ise anlamlı değildir. Bu durumun nedeni Yes/No yönteminde uzmanların sadece bir sınır grup öğrencisi düşünerek maddeler hakkında karar vermeleri olabilir.

Pearson Korelasyon Katsayısı ortalamalardan bağımsız bir hesaplama yaptığı için yöntemlere ait MGP'lerin ortalamaları bağımlı gruplar t testi ile karşılaştırılmıştır. Tablo 4'e göre Angoff ve Yes/No yöntemlerine ait MGP'lerin ortalamaları arasında manidar bir farklılık olmadığı; Angoff ve Ebel yöntemleri ile Yes/No ve Ebel yöntemlerine ait MGP'ler arasında ise 0,05 hata düzeyinde anlamlı bir farklılık olduğu görülmektedir.

Tablo 4. MGP Ortalamaları Arasındaki Farkın Testi

Karşılaştırılan Yöntemler	\bar{X}	t
Angoff Yöntemi	51,47	
Yes/No Yöntemi	57,94	-1,89
Angoff Yöntemi	51,47	
Ebel Yöntemi	47,12	2,13*
Yes/No Yöntemi	57,94	
Ebel Yöntemi	47,12	2,82*

*p<0.05

Analiz sonuçları Angoff ve Yes/No yöntemleri ile elde edilen MGP'ler arasında pozitif ve orta düzeyde bir ilişki olduğunu, iki yöntemde ait MGP'lerin ortalamaları arasında ise manidar bir farklılığın olmadığını göstermektedir. Bu iki bulguya göre Angoff ve Yes/No yöntemleri arasında MGP'ler bakımından uyum olduğu söylenebilir. Angoff ve Ebel yöntemleri ile elde edilen MGP'ler arasında pozitif ve yüksek düzeyde bir ilişki olduğu, iki yöntemde ait MGP'lerin ortalamaları arasında ise manidar bir farklılığın olduğu görülmektedir. Bu iki bulguya göre Angoff ve Ebel yöntemleri arasında MGP'ler bakımından uyum olduğu ancak MGP'lerin ortalamaları bakımından farklılıkların söz konusu olduğu söylenebilir. Yes/No ve Ebel yöntemleri ile elde edilen MGP'ler arasında ise anlamlı bir ilişkinin olmadığı ve iki yöntemde ait MGP'lerin ortalamaları arasında ise manidar bir farklılığın olduğu görülmektedir. Buna göre Yes/No ve Ebel yöntemleri arasında MGP'ler bakımından uyum olmadığı söylenebilir.

SONUÇLAR ve TARTIŞMA

Araştırma sonuçlarına göre üç yöntemde de uzman kararları arasında makul bir uyumun yakalandığı söylenebilir. Angoff, Yes/No ve Ebel yöntemleri ile belirlenen kesme puanlarının birbirinden farklı olduğu; kesme puanlarındaki bu farklılığın başarılı sayılan öğrenci sayısını ve yüzdesini etkilediği; kesme puanlarına göre başarılı sayılan öğrenci yüzdeleri arasında manidar bir farklılık olduğu ve öğrenci puanlarının 1-0 kodlanması durumunda ise yöntemlerin orta ve üstü düzeyde anlamlı bir uyum yakalayabildiği sonucuna ulaşılmıştır.

Elde edilen MGP'ler bakımından Angoff ve Ebel yöntemleri arasında yüksek derecede bir uyumun olduğu ancak MGP ortalamalarının manidar düzeyde farklılaştığı; Angoff ve Yes/No yöntemleri arasında ise orta derecede bir uyumun olduğu ve MGP ortalamalarının manidar düzeyde farklılaşmadığı görülürken; Yes/No yöntemi ve Ebel yöntemi arasında anlamlı bir uyumun olmadığı ve MGP ortalamalarının manidar düzeyde farklılaştığı sonucuna ulaşılmıştır. Angoff ve Yes/No yöntemlerinden elde edilen MGP'lerin birbiriyle uyumlu olması nedeniyle iki yöntemin beraber kullanılmasının zaman bakımından ekonomik olmayacağı durumlarda sadece biri seçilebilir. Bu üç yöntem arasında öğrencilerin başarı durumları bakımından belirli bir uyum olsa da başarılı sayılan öğrenci yüzdeleri bakımından manidar farklılıklar olduğu uygulamalarda göz önünde bulundurulmalıdır.

KAYNAKLAR

- Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*. Florida: Holt, Rinehart and Winston, Inc.
- Downing, S. M., Tekian, A. & 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 Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s15328015t1m1801_11#U496d_mPnns
- Ebel, R. L. (1972). *Essentials of educational measurement*. New Jersey: Prentice Hall.
- Hambleton, R. K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process, In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (ss. 89 – 116). New Jersey: Lawrence Erlbaum Associates.
- Impara, J. C. ve Plake, B. S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34(4), 353-366 Retrieved from <http://www.jstor.org/discover/10.2307/1435114?uid=2&uid=4&sid=21104259914143>
- 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.
- Murphy, K. R. & Davidshofer, C. O. (1991). *Psychological testing: Principles and applications*. New Jersey: Prantice Hall.
- Zieky, M. J. (2001). So much has changed: How the setting of cutscores has evolved since the 1980s, In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods and perspectives* (ss. 19-51). New Jersey: Lawrence Erlbaum Associates.

EXTENDED ABSTRACT

Introduction

In order to identify the students achievement or select students and personnel the psychological tests are used in Turkey as well as all over the World. Test results are important to make a decision about the individuals and making a decision requires some criteria that could be relative or absolute. Setting standard is a process that set the criterion/criteria to classify examinees into performance categories like master and non-master or successful-unsuccessful. In recent years several standard setting methods have been described in literature. In this study the test based standard setting methods which are Angoff method, Yes/No method and Ebel method were investigated. This methods based on judgemental decisions that determine how much is enough. Each judge (usually qualified expert in the content area) make the items decisions according to the standart setting methods and then minimum passing level (MPL) is calculated for each judges. The mean of the MPLs is the cut-off score for the test. In Angoff method judges are instructed to think of a group of “minimally acceptable” persons (borderline students) and for each item to estimate the proportion of the minimally acceptable group who would answer the item correctly. In Yes/No method the judges make a decision whether the borderline stundets answer the item correctly or not. In Ebel method the judges make two decisions which are the Levels of

Difficulty (Easy, Medium, Difficult) and Levels of Relevance (Essential, Important, Acceptable, Questionable) for the items.

There is no golden standard setting methods in literature and all of them based on the judgemental decisions which are subjective. Because its characteristics, the cut-off scores calculated with different standard setting methods may change for a test. The purpose of this study is to compare decisions and cut-off scores obtained from Angoff, Yes/No and Ebel standard setting methods to identify the relationship between this methods. For this aim it was investigated the judges decisions concordance; the differences between the percentages of students, considered as successful according to three different standard setting method; the compliance between the students' scores that have been tailored to binary as successful-unsuccessful according to three methods and the compliance between the cut-off scores of three methods

Method

This study is a associational and descriptive study because it aims of comparing three standard setting methods. The study groups were students and judges. The research data consisted of 489 student scores obtained from a achievement test of 20 items and 17 judges' decisions, made according to three different standard setting methods for an achievement test that was constructed by the researcher. The KR-20 reliability of test was 0.84. For the content validity (experts decisions) Kappa was calculated as 0.82. In the analysis of data, in order to set the compliance between judges, Kendall's coefficient of concordance (W) for Angoff and Ebel methods, and Cohen's Kappa statistics for Yes/No method have been calculated. Differences between the percentages of students, considered as successful according to three different standard setting methods, have been tested with the test of the difference between two correlated proportions. The students' scores have been tailored to binary as successful-unsuccessful according to three methods; and the compliance between them has been examined with Cohen's Kappa statistics. The compliance between the cut-off scores of three methods has been evaluated with Pearson Product-Moment Correlation Coefficient and the paired samples t test.

Results and Discussion

At the end of this research, W for Angoff method was calculated as 0,38 and for Ebel method as 0,32; Kappa for Yes/No method was calculated as 0,60. So it is seen that the compliance of judges in the methods was at a reasonable level. There is significant difference on 0,01 error level among percentages of students considered as successful according to these three different standard setting methods and there is a compliance between students' scores that have been tailored to binary as successful-unsuccessful according to three methods. It is concluded that there is a high relationship between Angoff and Yes/No methods; there is moderate relationship between Angoff and Ebel methods; there is no significant relationship between Yes/No and Ebel methods.