

**Kocaeli Üniversitesi**

**Eğitim Dergisi**

E-ISSN: 2636-8846

2020 | Cilt 3 | Sayı 2

Sayfa: 97-115



**Kocaeli University  
Journal of Education**


E-ISSN: 2636-8846

2020 | Volume 3 | Issue 2

Page: 97-115

ÜDS için Angoff ve madde haritalama yöntemleriyle  
standart belirleme

Setting standards for UDS with Angoff and item  
mapping methods

**Esin Bağcan Büyükturan,**  <https://orcid.org/0000-0001-8523-3951>

*Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, esinbagcanbuyukturan@ibu.edu.tr*

**R. Nükhet Çıkrıkçı,**  <https://orcid.org/0000-0001-8853-4733>

*İstanbul Aydın Üniversitesi, Eğitim Fakültesi, nukherdemirtasli@aydin.edu.tr*

Bu makale "Angoff ve Madde Tepki Kuramına dayalı madde haritalama yöntemleri ile ÜDS (üniversiteler arası kurul yabancı dil sınavı) için yeterli standartlarının belirlenmesi" isimli doktora tezinden üretilmiştir.

**ARAŞTIRMA MAKALESİ**

**Gönderim Tarihi**

*10 Eylül 2020*

**Düzeltilme Tarihi**

*10 Kasım 2020*

**Kabul Tarihi**

*10 Kasım 2020*

**Önerilen Atıf**

**Recommended Citation**

Bağcan Büyükturan, E., & Çıkrıkçı, R. N. (2020). ÜDS için Angoff ve madde haritalama yöntemleriyle standart belirleme. *Kocaeli Üniversitesi Eğitim Dergisi*, 3(2), 97-115. <http://doi.org/10.33400/kuje.793357>

## ÖZ

Bu çalışmada Angoff ve Madde Tepki Kuramına (MTK) dayalı Madde Haritalama yöntemleri ile belirlenen kesme puanlarının karşılaştırması amaçlanmıştır. Çalışma için ÖSYM tarafından 2000-2012 yılları arasında, yılda iki kez yapılan ve akademik kadrolara atanma aşamasında puanları ön koşul olarak kabul edilen ÜDS'nin 2005 yılı Sosyal Bilimler İngilizce testi arşiv verileri kullanılmıştır. Her iki yöntemin yürütülmesinde de yedi yargıcı ile çalışılmış ve bu yöntemlerle belirlenecek 3 kesme puanına ilişkin yeterlik düzeylerinin sınırında kalan adayların dil yeterlikleri Avrupa dil portfolyosu çerçevesi esas alınarak 15 kişilik bir uzman grubunun görüşleri doğrultusunda tanımlanmıştır. İki yönteme göre belirlenen kesme puanlarının "öğretim üyesi dışındaki öğretim elemanı kadrolarına atanabilme" düzeyi için anlamlı derecede farklı olduğu ancak "Doçentlik başvurusu yapabilme" ve "İngilizce okutmanlığı başvurusu yapabilme" düzeyleri için benzer olduğu bulunmuştur. MTK'ya dayalı belirlemelere bir geçerlik kanıtı oluşturmak amacıyla yanıtlayıcı örnekleme ait verilerden elde edilen madde parametrelerine göre maddeler yeniden haritalanmış ve yargıcılarla ikinci bir madde haritalama çalışması yürütülmüştür. Bu iki uygulamada verilen kararların birbirleri ile uyumlu olması, yapılan işlemin geçerliği için kanıt olarak kabul edilmiştir. Angoff ve MTK'ya dayalı Madde Haritalama yöntemleri ile üç ayrı kesme puanının belirlenmesinde yargıcılar arası uyum araştırılmıştır. MTK'ya dayalı Madde Haritalama yönteminde yargıcıların birbirleri ile uyumlu kararlar verdikleri, Angoff yöntemi için yargıcılar tarafından verilen kararların birbirleri ile uyumlu olmadığı görülmüştür. Yargıcıların iş yükünü azaltma ve onlara ortak standardı yakalama konusunda yol göstererek daha tutarlı sonuçlar elde etmeyi sağlama gibi avantajları sebebiyle MTK'ya dayalı Madde Haritalama Yöntemi'nin daha güvenilir ve kullanışlı olduğu görülmüştür.

*Anahtar Sözcükler:* Angoff yöntemi, kesme puanı, madde haritalama, madde tepki kuramı

## ABSTRACT

This study aims to compare the cutoff scores determined by Angoff and Item Response Theory (IRT) based Item Mapping methods. Archive data of UDS, which was made by ÖSYM was used in the study. Seven judges worked on the implementation of both methods. It has been found that the cutoff scores determined according to the two methods are significantly different for the level of "assignment to faculty members other than assistant professors, associate professors and professors", but similar for the levels of "applying for Associate Professorship" and "applying for English lecturer". In order to establish a proof of validity for the MTK based determinations, the items were re-mapped according to the item parameters obtained from the data of the respondent sample and a second item mapping study was conducted with the judges. The coherence of the decisions taken as a result of these two conduction are accepted as evidence for the validity of the applications. Consistency between judges was investigated in determining three different cut-off points with Angoff and MTK-based Item Mapping methods. Consistency between judges in determining three different cut-off points with Angoff and MTK-based Item Mapping methods was investigated. It was observed that judges made consistent decisions with each other in the Item Mapping method based on MTK, and the decisions made by the judges for the Angoff method were not consistent with each other.

*Keywords:* Angoff method, cut-off score, item mapping, item response theory

## GİRİŞ

Eğitim ve psikolojide birçok konuda karar alınırken ölçme sonuçlarının belli ölçütlerle karşılaştırılıp değerlendirilmesi yöntemine başvurulmaktadır. Crocker ve Algina (1986)'ya göre birçok durumda; örneğin, bir öğretim programını bitirip üst programa geçmek ya da bir uzmanlık alanında yeterliği ortaya koyan bir sertifika almak için adayların katıldıkları testlerde geçmeleri gereken belli bir ölçüt puan vardır. Bu amaçla kullanılan bir testte, test maddeleri bir performans alanını temsil ediyorsa, testi geliştiren bir kesme puanı (cut score / cutoff score) ya da standart belirlenmek istenir. Bu puanın belirlenmesi ile ilgili işlemler standart belirleme süreci olarak tanımlanır.

Standart belirleme (standard setting) kavramı, literatürde test maddelerinin doğru yanıtlanma olasılıklarından yola çıkarak bir testten alınacak puanları belirlenen kesme puanları ile belli düzeylere ayırma anlamında kullanılmaktadır (Irwin, 2007; Schulz, 2006; Shen, 2001; Zwick, Şentürk, & Wang, 2001).

Standart belirleme sürecinde ölçüt puan/kesme puanı, güçlüklerine göre dizilmiş maddeleri düzeylerine göre birbirinden ayırmak için kullanılır. Bu düzeyler geçti-kaldı olabileceği gibi acemi, çırak, uzman ve seçkin gibi düzeyler de olabilir (Dickinson, Thacker, Moody, & Hoffman, 2006). Berk (1986)'e göre seçilen standart belirleme uygulamasının pratik ve anlaşılır olması gerekir. Ayrıca belirlenen kesme puanlarının sadece test puanlarını değil bu puanlara karşılık gelen öğrenci performansını da yansıtması ve belirlenen kesme puanlarına ilişkin geçerlik kanıtlarının bulunması da önemlidir.

### Standart (Kesme Puanı) Belirleme Yöntemleri

Kesme puanları yargıcı kanısıyla belirlenebildiği gibi bazı görgül ve sistematik uygulamaları içeren çeşitli kesme puanı belirleme yöntemleri de bulunmaktadır. Nedelsky Yöntemi, Ebel Yöntemi, Jaeger Yöntemi, Sınır Grup Yöntemi, Zıt Gruplar Yöntemi, Angoff Yöntemi ve Madde Haritalama (MH) Yöntemi bunlardan bazılarıdır (Cizek, 2001).

Bu araştırmada test merkezli ve kolay anlaşılabilen bir yöntem olması sebebiyle seçilen Angoff Yönteminde; kabul edilebilir minimum kişinin doğru yanıtlayacağı düşünülen maddelere 1, diğerlerine 0 puan verilir. Madde puanları toplamı kesme puanını oluşturur. Yargıcılara sorulacak soruların formatı, yargıcıların birbirleri ile tartışarak mı yoksa bireysel olarak mı karar verecekleri gibi bazı ayrıntıları belirlenmemiş olan bu yöntem, araştırmacılar tarafından farklı biçimlerde uygulanmaktadır. Bu sebeple yöntem Angoff çeşitlemeleri olarak da adlandırılmaktadır (Cizek, 2001).

Jaeger (1989) yargıcı kanısına dayalı standart belirleme yöntemlerini test merkezli ve öğrenci/incelenen merkezli olmak üzere ikiye ayırmaktadır. Öğrenci merkezli standart belirleme yöntemleri standartları belirlenecek testi alan grubu tanımayı gerektirmektedir. Test merkezli yöntemlerde ise böyle bir gereklilik söz konusu değildir. Test merkezli yöntemler içinde en yaygın kullanılanı Angoff yöntemidir. Bunun sebebi yöntemin yargıcılar tarafından kolay anlaşılabilen, sade bir yöntem olmasıdır.

Yukarıda sıralanan standart belirleme yöntemleri yargıcı kanısına dayalı bir başka ifade ile yargısal yöntemlerdir. Bu yöntemlerle standart belirleme, yargıcıların "minimum yeterlikteki aday" ı doğru tanımladıkları ve bu yeterlik düzeyine göre madde güçlüklerini doğru tahmin ettikleri varsayımına dayanmaktadır (DeMauro, 1995; Wang, Wiser, & Newman, 2001).

Minimum yeterlikteki adayın, ya da her bir yeterlik düzeyindeki adayların yargıcılar tarafından farklı algılanmasının önüne geçilmesi için bazı yollar önerilmiştir. Bunlardan biri Cizek ve Bunch (2004)'in önerdiği gibi kesme puanlarının belirlenmesinde esas alınacak performans düzeyi tanımlarının, bu konudaki politikaları belirleyen ayrı bir uzman grubu tarafından yapılmasıdır. Bu tanımlar kesme puanları arasında kalan her bir kategorinin içerdiği bilgi ve becerilere ilişkin sözel tanımlardır ve standart belirleme sürecinde yargıcılara referans oluşturan ölçütler haline gelirler.

Sadece yargıcı kararlarına dayalı belirlemeleri esas almak dışında standart belirlemek için uygulama sonuçlarını dikkate alan istatistiksel yaklaşım da izlenebilir. Bu araştırmada konu edilen diğer yöntem olan Madde Tepki Kuramına (MTK) Dayalı Kesme Puanı Belirleme yöntemi istatistiksel yaklaşıma dayalı bir standart belirleme yöntemidir. MTK, temelleri 1920'li yıllarda ortaya çıkan ve 1980'li yıllarda eğitim ve psikoloji alanındaki çalışmalarda da kullanılmaya başlanan bir test kuramıdır. Bu kuram, klasik test kuramına göre avantajlı olan özellikleri sayesinde, klasik test kuramı ile çözülemeyen bazı ölçme problemlerine de çözüm getirmektedir (Hambleton & Swaminathan 1985; Thissen, 1991; Harvey & Hammer, 1999). Maddelerin haritalanarak standartların belirlenmesi sürecinde MTK'nın avantajları kullanılarak çözüm getirilen ölçme problemleri mevcuttur.

Madde Tepki Kuramına göre maddenin doğru yanıtlanma olasılığı örtük özelliğin (yetenek/yeterlik) bir fonksiyonudur. Buna göre kuramda yanıtlayıcıların özellikleri ile maddelerin özellikleri aynı ölçek üzerinde ölçeklenir (Crocker & Algina, 1986). Bu sayede belli bir yeterlik düzeyindeki bireyin henüz yanıtlanmadığı bir maddeyi doğru yanıtlanma olasılığı, her yetenek düzeyi için o maddenin doğru yanıtlanma olasılığını veren "madde karakteristik eğrisi" yoluyla kestirilir. Başka bir ifadeyle bir dizi maddedeki performanstan yola çıkarak başka bir dizi maddedeki performans kestirilebilir. Yargıcılara sunulmayan maddelere ilişkin kestirim yapılması imkanı sağladığından MTK'ya dayalı standart belirlemede yargıcılara maddelerin tamamı değil madde örneklemini sunularak az madde ile kestirim yapılabilir ve zamandan tasarruf edilir. Ayrıca kimlerin hangi puanı alacağı önceden kestirilebildiği için yargıcılar belirledikleri kesme puanına karşılık gelen geçme oranlarını da bileceklerdir. Örneğin; "65 puanı kesme puanı olarak kabul edersek grubun %84'ü geçer" (Hambleton, Swaminathan, & Rogers, 1996; Wang vd., 2001; Ryan, 2003).

Madde tepki kuramına dayalı kesme puanı belirlemek için izlenen en yaygın yol maddelerin MTK'ya göre kestirilen güçlük düzeylerini gösteren, madde haritası denen ve yargıcılara kesme puanı belirlemede yol gösteren bir histogram grafik oluşturulması; yargıcıların madde haritaları üzerinde, minimum yeterlik düzeyindeki bireyin %50 olasılıkla doğru yanıtlayabileceğine inandıkları en zor maddenin hemen altına bir işaret koymasındadır. Böylece her yargıcı tarafından bir kesme noktası belirlenmiş olur. Bu süreçte yargıcılara madde haritasında yer alan maddeler, MTK'ya göre belirlenmiş güçlük düzeylerine göre kolaydan zora doğru sıralanarak kitapçık biçiminde verilir ve bu kitapçığa Sıralanmış Madde Kitapçığı (SMK) denir. SMK, yargıcıların; maddelerin göreceli güçlüklerini anlamalarına yardımcı olur. Nihai kesme noktası olarak belirlenen maddenin ham puanı, nihai kesme puanı anlamına gelir. Maddeler kolaydan zora doğru sıralanmış olduğundan bu maddenin ham puanı MTK'ya dayalı belirlemelere göre "kendisinden daha kolay olan" maddelerin ham puanları toplamı kadardır. (Wang vd., 2001; Cizek & Bunch, 2004; Karantonis & Sireci, 2006; Lissitz & Kroopnick, 2007).

Araştırmalarda MTK'ya dayalı madde parametrelerinin belirlenmesi aşamasında bir, iki ya da üç parametrelili modellerin kullanılabildiği, doğru yanıtlanma olasılıklarının %50, %67 ya da %80 olarak kabul edilebildiği, ham kesme puanları yanında maddelerin yeterlik (theta,  $\theta$ ) değerlerine dayalı kesme puanlarının da verilebildiği, nihai kesme puanı belirlenirken yargıcıların belirlediği kesme puanlarının ortancasının ya da yeterlik değerlerinin ortalamasının alınabildiği görülmektedir (Zwick vd., 2001; Cizek & Bunch, 2004; Seçil, 2009).

### **Kesme Puanlarının İsabetliliği**

Eğitimde sıklıkla üzerinde kararlar verilen yeterlik alanlarından biri de yabancı dildeki yeterliklerdir. Bachman (1990), belirlenen kesme puanlarına dayanarak yapılan yeterlik sınıflamalarında iki türlü sınıflama hatasından söz etmiştir. Gerçekte puanı kesme puanının altında kalması gereken bir adayı "yeterli" olarak sınıflamak yanlış pozitif, gerçekte puanı kesme puanının üstünde olması gereken adayı "yetersiz" olarak sınıflamak ise yanlış negatif sınıflama hatasıdır. Örneğin bir sonraki dil kuruna geçenleri belirlemek için uygulanan bir testte yapılan yanlış pozitif sınıflama hatası sonucu öğrenci hazır olmadığı bir kurla karşılaşacak, derslere ilgi ve motivasyonu düşecek, bu durum başarısızlığa sebep olacaktır. Yanlış negatif bir sınıflama

hatası yapıldığı durumda ise öğrenciye onun öğrenmelerini geliştirmeyecek, gereksiz bir ek eğitim verilmiş olacaktır. Bu sebeple isabetli sınıflamalar yapabilmek için en uygun kesme puanının belirlenmesi gerekir.

İsabetli kesme puanları elde etmek için sistematik standart belirleme yollarını izlemenin yanı sıra bu kesme puanlarının geçerliğine ve güvenilirliğine ilişkin kanıtların da elde edilmesi gerekir. Camilli, Cizek ve Lugg (2001), standart belirleme çalışmasında hedef gruptan seçkisiz olarak seçilecek bir grubun kesme puanları ile tüm gruba ilişkin kesme puanlarının benzer olmasını, belirlenen kesme puanlarının geçerlik kanıtı olarak kabul etmektedir. Bachman (1990) da kesme puanlarının güvenilirlik kanıtı olarak, her bir yargıcının her bir maddeyi yerleştirdiği düzeye bakarak yargıcı değerlendirmeleri arasındaki uyuma ilişkin bir katsayı hesaplanabileceğini belirtmiştir.

Harch, Pant ve Köller (2010)'e göre belirlenen kesme puanlarının geçerli ve güvenilir olmasına ilişkin kanıtlar elde edilmesinin dışında kesme puanlarının inandırıcı ve savunulabilir olması için belirlemeleri yapan yargıcıların sadece test kapsamını öğreten eğitimcilerden oluşması uygun değildir. Yargıcılar arasında bu testin puanlarının kullanılması aşamasında yer alan ve testi alan grup konusunda fikri olan uzmanların da bulunması gerekir.

Bu konu, standart testlerde karşılanması gereken özellikler kapsamında da ele alınmıştır. Eğitim ve psikoloji alanlarında kullanılan testlere ilişkin olarak geliştirilen standartlarından biri (4.20 nolu standart) kesme puanlarını belirlerken test performansı ile ölçüt arasında ampirik veriden yararlanarak ilişki kurulmasının gerekli olduğunu vurgulamaktadır. Profesyonel kararlar verilebilmesi için en uygun standart belirleme yaklaşımının seçilmesi gerekmektedir. Bir başka standart ise (4.21 nolu standart) yeterli düzeylerini ya da geçme kalma kararlarını belirleyen kesme puanlarının sadece yargısal yolla belirlendiği durumlarda, yargıcılara görevlerini tam anlamalarını ve doğru karar vermelerini sağlayacak sistematik bir yol çizilmesinin gerekli olduğunu belirten standarttır. Ayrıca (14.14 nolu standart) yeterli belirleme ve belgeleme testlerinde, değerlendirilen bilgi ve beceriler açıkça tanımlanmalı, bunların "yeterli performans" için ne açıdan gerekli olduğu açıklanmalıdır (AERA, APA, & NCME, 2002).

Moy (1981)'a göre dil yeterliklerini belirleyen testlerde standart belirlemeye ilişkin en büyük problem; standartların, bu testi alan gruba bağlı başka bir ifade ile norm dayanaklı olarak belirlenmesi ya da kesme puanlarının gelişigüzel kabul edilmesidir. Türkiye'de, ÖSYM tarafından yapılan ÜDS, akademik kadrolara atanmak ve lisansüstü eğitim almak için geçilmesi gereken ön koşul niteliğindeki bir yabancı dil sınavıdır. ÜDS'nin farklı karar alanları için kesme puanları sabittir. Farklı amaçlar için alınması gereken bu minimum puanlar şöyledir; öğretim üyesi dışındaki öğretim elemanı kadrolarına naklen veya açıktan yapılan her türlü atamada ÜDS'den en az 50 puan, doçentlik başvurusu yapabilmek için en az 65 puan, yabancı dil okutmanı başvurularında en az 80 puan almak gerekmektedir (ÖSYM, 2012).

Görüldüğü gibi ÜDS puanlarına ilişkin verilen kararlar mutlak ölçütlere dayanmaktadır. Bu çalışmaya ÜDS'nin konu edilmesinin en önemli sebebi bu sınavda ölçme sonuçları kullanılarak mutlak değerlendirme yapılmasıdır. Bu doğrultuda bu çalışmada ÜDS Sosyal Bilimler İngilizce Testi için Angoff ve MTK'ya dayalı Madde Haritalama yöntemleri ile yeni mutlak ölçütler başka bir ifade ile kesme puanları belirlemek amaçlanmıştır. Bu genel amaç doğrultusunda ÜDS Sosyal Bilimler İngilizce testi puanlarının yorumlanmasında "öğretim elemanı olarak atanabilme", "doçentlik başvurusu yapabilmeye" ve "yabancı dil okutmanlığı başvurusu yapabilmeye" yeterlik düzeyleri için, Angoff ve MTK'ya dayalı Madde Haritalama Yöntemleri ile belirlenen kesme puanları belirlenmiştir. Bir geçerlik kanıtı olarak, tüm gruptan random olarak seçilen bir yanıtlayıcı örnekleme ve tüm grup için MTK'ya dayalı Madde Haritalama yöntemiyle belirlenen kesme puanları arasında fark olup olmadığı araştırılmıştır. Bir güvenilirlik kanıtı olarak ise, kesme puanlarının belirlenmesinde Angoff ve MTK'ya dayalı Madde Haritalama yöntemleri kullanıldığında yargıcılar arası uyum katsayıları hesaplanmıştır. Ayrıca Angoff ve MTK'ya dayalı Madde Haritalama Yöntemleri ile belirlenen kesme puanları arasında fark da araştırılmıştır.

## YÖNTEM

Bu çalışma, var olan duruma müdahale etmeden değişkenler arasındaki birlikte değişimin derecesini inceleyen ilişkisel tarama modelinde bir araştırmadır.

### Katılımcı Bilgileri

Bu araştırmanın çalışma grubunu ÖSYM'den elde edilen veriler doğrultusunda, 2005 yılı ÜDS Sosyal Bilimler İngilizce Testini alan "Eğitim Bilimleri ve Öğretmen Yetiştirme (Sosyal bölümler) ile Beşeri ve İdari Bilimler alanlarındaki tüm katılımcılar olan 3731 yanıt veri seti oluşturmaktadır. Bu yanıtlayıcıların 1993'ü (%53,42) erkek ve 1738'i (%46,58) kızdır.

Bu araştırmada İngilizce okutmanlığı başvurusu yapmaya hak kazanacak, doçentlik başvurusu yapmaya hak kazanacak ve öğretim üyesi dışındaki öğretim elemanı kadrolarına başvurmaya hak kazanacak İngilizce açısından minimum yeterlikteki üç aday grubunun okuduğunu anlama boyutundaki dil yeterliklerinin tanımlanması sağlanmıştır. Bu tanımlamalar Avrupa ortak dil çerçevelerinde "okuma" boyutunda farklı düzeylere karşılık gelen beceriler kapsamında, Ankara Üniversitesi Eğitim Bilimleri Fakültesi'nde görev yapan 10 profesör ve Ankara Üniversitesi Rektörlüğü'ne bağlı olarak çalışan 5 İngilizce okutmanından görüş alınarak yapılmıştır.

Ayrıca araştırma soruları kapsamında, 2005 yılı ÜDS Sosyal Bilimler İngilizce Testi'nin standartlarını Angoff ve MTK'ya dayalı biçimde belirleme sürecinde Üniversitelerin İngilizce öğretmenliği ile İngiliz Dili Edebiyatı bölümlerinde görev yapan yedi öğretim üyesi yargıcı olarak çalışmaya katkı getirmiştir.

### Veriler ve Toplaması

Araştırma, 2005 yılı bahar döneminde uygulanmış ÜDS Sosyal Bilimler İngilizce Testi maddeleri ve yanıtlayıcı verileri kullanılarak gerçekleştirilmiştir. Kesme puanlarının belirlenmesinde MTK'ya dayalı MH yöntemi ve yargıcı kanısına dayalı Angoff yöntemi kullanılmıştır. Araştırmada yargıcı kararlarına dayalı kesme puanı belirleme yollarından sadece Angoff Yönteminin kullanılmasının sebebi; bu yöntemin testi alan adayların yargıcılar tarafından tanınmasını gerektirmeyen test merkezli bir yöntem olmasıdır. Ayrıca Angoff yöntemi daha kesme puanı belirleme çalışması yapmamış yargıcılar için anlaşılması kolay bir kesme puanı belirleme yöntemidir.

Çalışmada, kesme puanları belirleme işlemlerine geçilmeden önce, 3 uzmanla kesme puanı belirleme çalışmasının ön denemesi yapılmıştır. Bu deneme uygulaması sonunda, kesme puanı belirlemede esas alınacak yeterlik düzeylerinin sınırında bulunan "İngilizce okutmanlığı başvurusu yapmaya hak kazanacak, doçentlik başvurusu yapmaya hak kazanacak ve öğretim üyesi dışındaki öğretim elemanı kadrolarına başvurmaya hak kazanacak İngilizce açısından minimum yeterlikteki adayların" yargıcılar tarafından algılanmasında güçlükler ve yargıcılar arası farklılıklar olduğu gözlenmiştir. Bu sebeple söz konusu adayların tanımlanması gerektiğine karar verilmiştir.

Bunun için Cizek ve Bunch (2004)'un önerdiği gibi yargıcı grubundan bağımsız bir uzman grubu ile çalışılmıştır. Çalışmada Avrupa Dil Portfolyosu'ndaki okuduğunu anlama becerilerini içeren bir form kullanılmıştır. Formda yer alan düzeyler ve bu düzeylere ilişkin tanımlar Avrupa Dil Portfolyosu resmi web sitesinden (adp.meb.gov.tr, 2011) alınmıştır. Formlar 10 profesör ve beş okutman olmak üzere 15 uzmana verilmiş, uzmanlardan; İngilizce okutmanlığı, doçentlik ve üniversitede öğretim elemanı kadrolarına başvurmaya hak kazanacak İngilizce açısından minimum yeterlikteki adayı temsil ettiğini düşündükleri tanımlamayı işaretlemeleri istenmiştir. Her düzeyin minimum yeterlik sınırında bulunan bu üç aday için en çok (8/15) işaretlenen tanım, o adayın tanımı olarak kabul edilmiştir.

Bu işlemin gerçekleştirilmesinde uzmanlar arasındaki uyum araştırılmış bunun için uyum katsayısı Fleiss'in Kappa formülü ile hesaplanmış ve 0,55 olarak bulunmuştur. Bu katsayı

yargıcılar arasında çoğunlukla uyum olduğunu göstermektedir (Gwet, 2001). İşlem sonunda ortaya çıkan tanımlar şöyledir;

İngilizce okutmanlığı başvurusu yapmaya hak kazanacak İngilizce açısından minimum yeterlikteki aday; kullanım kılavuzları, uzmanlık alanına yönelik makaleler ve yazınsal yapıtlar gibi soyut, yapısal ve dil bilgisel açıdan karmaşık hemen hemen tüm metin türlerini kolaylıkla okuyabilir ve anlayabilir. Doçentlik başvurusu yapmaya hak kazanacak, İngilizce açısından minimum yeterlikteki aday; yazarların belirli tutum ya da görüşü benimsedikleri, güncel sorunlarla ilgili makaleleri ve raporları okuyabilir. Çağdaş edebi düzyazıyı anlayabilir. Öğretim üyesi dışındaki öğretim elemanı kadrolarına başvurmaya hak kazanacak İngilizce açısından minimum yeterlikteki aday; meslekle ilgili ya da günlük dilde en sık kullanılan sözcükleri içeren metinleri anlayabilir. Kişisel mektuplarda belirtilen olay, duygu ve dilekleri anlayabilir. Bu tanımlar, kesme puanlarının oluşturulmasında esas alınacak üç adayı, ilgili beceri bakımından daha somut ve anlaşılır kılmak için, MTK'ya dayalı MH ve Angoff yöntemine göre kesme puanı belirleme sürecinde yargıcılara verilen yönergelere eklenmiştir.

Asıl uygulamada, Angoff yöntemine göre kesme puanı belirlenirken yargıcılara ÜDS 2005 Sosyal Bilimler için İngilizce testindeki 80 madde, her kartta bir madde yer alacak şekilde madde kartları biçiminde verilmiştir.

Angoff yönteminin ayrıntıları belirlenmediğinden bu yöntemde yargıcılara sorulacak soruların formatı da araştırmacılar tarafından belirlenmiştir (Cizek, 2001). Bu çalışmada yargıcılara minimum yeterlikteki bireyin maddeyi doğru yanıtlama olasılığı "oran" cinsinden sorulmamış, "doğru yanıtlayabilir mi?" sorusu sorularak Impara ve Plake (1997)'in belirttiği gibi yargıcılar tarafından daha kolay anlaşılabilir "evet/hayır" yöntemi izlenmiştir. Buna göre her bir madde kartının altında sırasıyla şu sorular yer almaktadır;

1. İngilizce okutmanlığı başvurusu yapmaya hak kazanacak, İngilizce açısından minimum yeterlikteki adayı düşünün. Sizce bu aday, bu maddeyi doğru yanıtlayabilir mi?
2. Doçentlik başvurusu yapmaya hak kazanacak, İngilizce açısından minimum yeterlikteki adayı düşünün. Sizce bu aday, bu maddeyi doğru yanıtlayabilir mi?
3. Öğretim üyesi dışındaki öğretim elemanı kadrolarına başvurmaya hak kazanacak, İngilizce açısından minimum yeterlikteki adayı düşünün. Sizce bu aday, bu maddeyi doğru yanıtlayabilir mi?

Yargıcılar, yukarıda verilen üç ölçüt soru kapsamında maddelere ilişkin incelemelerini bireysel olarak ortalama 2 saatlik bir sürede yürütmüşlerdir. Yargıcılar bu soruları "evet" ve "hayır" seçeneklerinden birini işaretleyerek cevaplamışlardır. ÖSYM tarafından ortaya konan üç amaca yönelik minimum puan sıralaması temel alınarak bu üç sorunun aşamalı biçimde yanıtlanması sağlanmıştır. Buna göre, her bir madde kartı için, bir soruda hayır cevabı veren yargıcının daha alt düzeyde yer alan adaylar için takip eden diğer soruları yanıtlamasına gerek olmadığı, sonraki soruya/sorulara da hayır cevabı vermiş olarak kabul edileceği yargıcılara hem yazılı yönerge yolu ile hem de sözlü olarak bildirilmiştir. Böylece yargıcı bir madde kartının altında yer alan sorulardan ilkinde hayır cevabı verirse ikinci soruya geçmemiş, benzer biçimde ikinci soruya hayır cevabı veren yargıcı üçüncü soruyu yanıtlamamıştır. Süreç sonunda yargıcıların tamamının işaretlemelerini bu kurala uygun biçimde yaptıkları görülmüştür.

Yargıcıların maddeleri hatırlamasını önlemek amacıyla Angoff ve MTK'ya dayalı madde haritalama çalışması 10 gün arayla uygulanmıştır. MTK'ya dayalı madde haritalama sürecinde yargıcıların iş yükünü azaltmak için, MTK'nın bir dizi maddeyi haritalayarak diğerleri hakkında yorum yapmaya olanak vermesi özelliğinden yararlanılarak 80 maddeyi temsil edecek 16 madde seçilmiş ve harita bu 16 madde ile oluşturulmuştur. Bu 16 temsilci maddenin belirlenmesi için 2 parametrelili modele göre parametreleri kestirilmiş olan maddeler b değerlerine göre küçükten büyüğe doğru sıralanmış, maddelerin b değerleri standart z puanlarına dönüştürülmüş ve her "1 z"lik aralıklar tabaka gibi kabul edilerek her aralıktan (tabakadan) özet madde haritasında yer almak üzere madde sayısının 1/5'i oranında temsilci madde seçilmiştir.

Bu maddeler b değerlerine göre kolaydan zora doğru sıralanarak her sayfada bir madde olacak şekilde SMK'na dönüştürülmüştür ve bu 16 maddenin b değerleri esas alınarak özet madde haritası oluşturulmuştur. Daha sonra araştırma soruları kapsamında, geçerlik kanıtı elde etmek amacıyla tüm yanıtlayıcıların %50'sini oluşturan 1868 kişilik bir örneklem grup seçkisiz olarak seçilmiş ve bu yanıtlayıcı örneğine ait verilerden madde parametreleri tekrar kestirilmiştir. Aynı işlem bu verilerden elde edilen b değerlerine de uygulanarak 80 maddenin 16'sı temsilci madde olarak seçilmiştir. Bu maddelerin seçiminde de tabakalı örnekleme yöntemi kullanılmıştır. Temsilci maddelerin aynı madde havuzundan seçkisiz olarak seçilmesi sebebiyle üç madde her iki madde haritasında da ortak maddeler olarak yer almış ve seçkisizliği bozmamak adına bu maddeler değiştirilmemiştir. Bu 16 madde de benzer biçimde "sıralanmış madde kitapçığı"nda yer almış ve maddelerin b değerlerine göre ikinci bir özet madde haritası oluşturulmuştur (Ek 5). Bu sıralanmış madde kitapçıkları ve madde haritaları kullanılarak, yargıcılarla MTK'ya dayalı, birbirinden bağımsız iki ayrı standart belirleme çalışması yürütülmüştür.

Tüm gruba ilişkin b değerlerine dayalı madde haritalama çalışması yürütülürken yargıcılardan sıralanmış madde kitapçıklarındaki maddelere ve bunların madde haritasındaki yerine bakarak, kesme puanı belirlemede esas alınacak yeterlik düzeylerinin sınırında bulunan adayların %50 olasılıkla doğru yanıtlayacağını düşündükleri en zor maddelerin hemen üstüne bir işaret koymaları istenmiştir. Yargıcılar SMK ve madde haritasında yer aldığı gibi, maddeleri kolaydan zora doğru incelemişlerdir. Cizek ve Bunch (2004)'a göre SMK kullanırken yargıcıların bilmesi gereken en önemli şey her maddenin bir önceki maddeden daha zor olduğudur. Bunun için yargıcılar öncelikle "öğretim üyesi dışındaki öğretim elemanı kadrolarına başvurmaya hak kazanacak" daha sonra "doçentlik başvurusu yapmaya hak kazanacak" ve son olarak da "İngilizce okutmanlığı başvurusu yapmaya hak kazanacak" İngilizce açısından minimum yeterlikteki adayın %50 olasılıkla doğru yanıtlayacağını düşündükleri en zor maddenin hemen üstüne bir işaret koymuşlardır. Yargıcılar bu çalışmayı araştırmacı eşliğinde yürütmüş ve bu süreçte yukarıda sıralanan üç adayın okuduğunu anlama becerileri ile ilgili özelliklerini içeren daha önce oluşturulmuş tanımları da yargıcılara verilmiştir. Bu süreç yürütülürken yargıcıların Angoff çalışmasında maddelere ilişkin verdikleri kararları görmelerine ya da önceden verdikleri bu kararları değiştirmelerine izin verilmemiş, iki çalışma birbirinden bağımsız yürütülmüştür. İlk madde haritalama çalışmasından sonra aynı işlem yanıtlayıcı örneğine ait verilerden elde edilen madde haritası ve sıralanmış madde kitapçığı üzerinde de yürütülmüştür.

### Verilerin Analizi ve Yorumlanması

Madde Tepki Kuramı'nın avantajlarından yararlanabilmek için verilerin bu kuramın varsayımlarını karşılaması gerekmektedir. MTK'nın "testin tek boyutluluğu, madde parametrelerinin değişmezliği, yetenek kestirimlerinin değişmezliği, yerel bağımsızlık ve hız testi olmama" varsayımları vardır (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985). ÜDS Sosyal Bilimler İngilizce Testi'ne ait verilerin bu kuramın varsayımlarını karşıladığı tespit edilmiştir.

Angoff yöntemine göre yürütülmüş olan işlemde yargıcılardan alınan cevaplar doğrultusunda her bir maddenin üç kesme puanından, hangisinin altında ya da üstünde kaldığı belirlenmiştir. Bir maddenin, belli bir kesme puanının üstünde kalması oy çokluğu koşulu ile belirlenmiştir. Bunun için yedi uzmanın en az dördünden ilgili kesme puanına ilişkin madde kartının altında yer alan soru için "evet" cevabı alınmış olması koşulu aranmıştır.

Buna göre 1.kesme puanı şöyle belirlenmiştir; "Öğretim üyesi dışındaki öğretim elemanı kadrolarına başvurmaya hak kazanacak İngilizce açısından minimum yeterlikteki aday" ın oy çokluğu ile doğru yanıtlayacağı düşünülen maddelere "1", diğerlerine "0" puan verilmiştir. Bu puanlar maddelerin nihai ham puanları olmuş ve bu puanlar toplanarak kesme puanı belirlenmiştir. İkinci ve üçüncü kesme puanlarının belirlenmesinde de aynı yol izlenmiştir.



MTK'ya dayalı MH yönteminde ise ilk kesme puanı belirlenirken yargıcıların her biri madde haritasında "Öğretim üyesi dışındaki öğretim elemanı kadrolarına başvurmaya hak kazanacak İngilizce açısından minimum yeterlikteki aday" ın %50 olasılıkla doğru yanıtlayabileceğini düşündükleri en zor maddeyi işaretlemişlerdir. Bu kesme puanı için yedi yargıcı tarafından yapılmış yedi ayrı işaretleme olduğundan, işaretli yedi maddenin b değerlerinin ortancası alınmış, buna karşılık gelen maddenin madde haritasındaki yeri nihai kesme noktası olarak kabul edilmiştir. Bu durumda b parametreleri esas alınarak kolaydan zora doğru sıralanmış olan maddelerden kesme noktası olarak belirlenen maddenin altında kalan maddelerin sayısı ham kesme puanı olarak kabul edilmiştir. İkinci ve üçüncü kesme puanları için de yargıcılar tarafından bu düzeylerin sınırında yer alan adaylar düşünülerek işaretlenen maddeler yoluyla aynı yöntem izlenerek kesme puanları hesaplanmıştır. Yanıtlayıcı örnekleme için verilerden elde edilen b değerleri esas alınarak oluşturulmuş madde haritası üzerinde de aynı işlem yürütülmüştür.

Angoff yöntemine göre elde edilen kesme puanları ham puan (madde sayısı) ve 100 puan üzerinden hesaplanan kesme puanı olarak ifade edilirken MTK'ya dayalı MH yöntemine göre elde edilen kesme puanları bunların yanında b değerleri ve standart puan cinsinden de ifade edilmiştir. Maddelerin kestirilen b değerleri -2,05 ile 6,18 arasında değişmektedir, b değerleri ortalaması 0,85; standart sapması 1,45 olarak hesaplanmıştır. Kesme puanlarına karşılık gelen b değerlerini daha anlaşılabilir kılmak için bu değerler, ortalaması 250 ve standart sapması 50 olan standart puan ölçeğine dönüştürülmüştür.

Bir geçerlik kanıtı olarak, yanıtlayıcı örnekleme için verilerden elde edilen 7 ayrı kesme puanı ile tüm gruptan aynı yolla belirlenen 7 kesme puanının uyumunu araştırmak amacıyla ilişkili iki puan dizisi arasındaki farkın manidarlığını test etmek için kullanılan Wilcoxon işaretli sıralar testi yapılmıştır. Burada sözü edilen kesme puanı, kesme noktası olarak belirlenen maddelerin "b" değerleridir. Yargıcılar tarafından iki harita ile belirlenen 1., 2. ve 3. kesme puanları arasındaki farkın manidarlığı araştırmak için üç ayrı karşılaştırma yapılmıştır.

Angoff yönteminde yargıcılara her bir madde kartının altında 3 ayrı yeterlik düzeyinin sınırında kalan aday için "bu aday bu soruyu doğru yanıtlayabilir mi?" sorusu sorulmuş ve yargıcılardan "evet-hayır" biçiminde cevaplar alınmıştır. Alınan "evet" cevapları "1", hayır cevapları "0" biçiminde iki kategorili olarak kodlanmıştır. Üç ayrı kesme puanı için yargıcıların belirlemelerine ilişkin üç ayrı uyum indeksi hesaplanmıştır. Uyum indeksinin hesaplanmasında 1-0 biçimindeki verilere ve ikiden fazla yargıcı değerlendirmesine uygun olan Cochran Q testi tercih edilmiştir.

MTK'ya dayalı MH yönteminde de bir madde; yargıcının, birinci kesme puanını belirlemek üzere koyduğu işaretin altında kalıyorsa bu maddeye 1, üstünde yer alıyorsa 0 kodu verilmiştir. İkinci ve üçüncü kesme puanları için de aynı kodlama yapılmıştır. Üç ayrı kesme puanı belirleme işlemi için yargıcılar arası uyumu gösteren üç ayrı uyum indeksi hesaplanmıştır. Angoff yöntemindekine benzer biçimde yargıcılar arası uyumun hesaplanmasında Cochran Q testi tercih edilmiştir. Yargıcıların madde haritalarını kullanarak belirleme yaparken 80 madde yerine 16 maddeyi değerlendirmiş olmaları sebebiyle bu işlem için yargıcılar arası uyum hesaplanırken sadece 16 madde için yapılan incelemelere ait veriler kullanılmıştır.

Üç kesme noktasından her biri için ilgili kesme noktasının altında kalan maddelerin sayısı aynı zamanda ham kesme puanına karşılık gelmektedir. Benzer biçimde ilgili kesme noktasının altında kalan maddelerin yüzdesi de 100 üzerinden hesaplanan kesme puanına karşılık gelmektedir. Bu durumda Angoff ve MTK'ya dayalı madde haritalama yöntemlerine göre belirlenen kesme puanları arasındaki farkın manidarlığını araştırmak için kesme noktalarının altında kalan maddelerin yüzdesinin karşılaştırılması yoluna gidilmiştir. Bu karşılaştırma "ilişkili iki yüzde arasındaki farkın manidarlığı için z testi" ile yapılmıştır. Bunun için üç kesme puanından her biri için bu test ayrı ayrı uygulanmıştır.

Araştırmada, her bir maddenin Angoff ve MTK'ya dayalı MH yöntemlerine göre ilgili kesme puanının altında ya da üstünde kalma durumunu gösteren, iki kategorili gözlem çiftlerinden

oluşan veriler söz konusudur. 1-0 biçiminde kodlanan veriler için çift gözlemler 2x2'lik bir kontincensi tablosunda gösterilmiştir.

### Şekil 1

*Verilerin Kontincensi Tablosunda Gösterilmesi*

		Frekanslar		
		Gözlem 2	0	
Gözlem 1	1	B	A	A+B
	0	D	C	C+D
		B+D	A+C	N

Oluşturulan tablodan yararlanılarak Formül 1'e göre kritik oran z değeri hesaplanmıştır. Kritik oran z değerinin 1.96'dan büyük olması iki yüzde arasındaki farkın .05 düzeyinde manidar sayılmasını gerektirir (Akhun, 1982);

### Formül 1

*Kritik Oran Z Değerinin Hesaplanması*

$$Z = \frac{B - C}{\sqrt{B+C}}$$

Bu araştırmada "A" her iki yöntemle göre, ilgili kesme noktasının üstünde yer alan, "B" Angoff yöntemine göre ilgili kesme noktasının altında yer alırken MTK'ya dayalı MH yöntemine göre bu noktanın üstünde yer alan maddelerin sayısını göstermektedir. Tablodaki "C" hücresi Angoff yöntemine göre ilgili kesme noktasının üstünde kalırken MTK'ya dayalı MH yöntemine göre bu noktanın altında kalan, "D" ise her iki yöntemle göre de bu noktanın altında yer alan maddelerin sayısını göstermektedir. Buna göre, iki standart belirleme yöntemi ile belirlenen 3 kesme puanının altında ve üstünde kalan madde sayıları 2X2'lik üç ayrı kontincensi tablosunda gösterilmiş, kullanılan yöntem değiştiğinde 1., 2. ve 3. kesme puanlarının altında ve üstünde kalan madde yüzdelerinin değişip değişmediği araştırılmıştır. Araştırmanın verileri SPSS, BILOG-MG, ve Microsoft Excel programları ile, .05 manidarlık düzeyinde analiz edilmiştir.

### Araştırma Etiği

Bu araştırmanın planlanmasından, uygulanmasına, verilerin toplanmasından verilerin analizine kadar olan tüm süreçte "Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi" kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan "Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler" başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Bu çalışmanın yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulmuş; toplanan veriler üzerinde herhangi bir tahrifat yapılmamış ve bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir.

### BULGULAR

Bu bölümde araştırma sorularına ilişkin bulgular yer almaktadır. Bu doğrultuda Angoff yöntemine göre belirlenen kesme puanları Tablo 1'de verilmiştir.

**Tablo 1***Angoff Yöntemine Göre Belirlenen Kesme Puanları*

	Madde sayısı	Puan (100 puan üzerinden)
1.kesme puanı	23	28,75
2.kesme puanı	51	63,75
3.kesme puanı	77	96,25

MTK'ya dayalı MH yöntemine göre belirlenen kesme puanları ve kesme noktasındaki maddelerin b değerleri ile bu değerlerin ortalaması 250 standart sapması 50 olan standart puan cinsinden ifadesi Tablo 2'de verilmiştir.

**Tablo 2***MTK'ya Dayalı MH Yöntemine Göre Belirlenen Kesme Puanları*

	Madde Sayısı	Puan (100 puan üzerinden)	Kesme noktasındaki maddenin b değeri	Standart puan
1.kesme puanı	37	46,25	0,54	239
2.kesme puanı	49	61,25	1,09	258
3.kesme puanı	77	96,25	3,75	350

MTK'ya dayalı MH Yöntemi ile belirlenen kesme puanlarının geçerliğine ilişkin bulguya ulaşmak için öncelikle geçerlik ölçütünü oluşturacak madde haritalama işlemine ilişkin güvenilirlik kanıtları aşağıda sunulmuştur. Bu amaçla yargıcılar tarafından verilen kararlar açısından yargıcılar arası uyumun araştırılması için yapılan Cochran Q testi sonuçları Tablo 3'te gösterilmektedir.

**Tablo 3***Yanıtlayıcı Örneklemine Ait Veriler Kapsamında Yapılan Madde Haritalama Çalışmasında Yargıcılar Arası Uyumunu Gösteren Cochran Q Testi Sonuçları*

	1.kesme puanı için	2.kesme puanı için	3.kesme puanı için
Cochran Q	8,67	13,29*	12,00

Not: \* $p < .05$ ,  $N=16$ ,  $Sd=6$

Tablo 3'e göre 1. ve 3. kesme puanının belirlenmesinde yargıcıların verdikleri kararlar arasında manidar bir fark ( $Q_1=8,67$ ,  $Q_3=12$ ,  $p > .05$ ) olmadığı bulunmuştur. Bu durum yanıtlayıcı örneklemeden elde edilen verilere göre hesaplanan b değerleri çerçevesinde yürütülen madde haritalama çalışmasında yargıcıların 1. ve 3. kesme puanının belirlenmesi konusunda uyum gösterdikleri biçiminde özetlenebilir. Aynı madde haritalama çalışmasında belirlenen ikinci kesme puanı için ise yargıcı kararları arasındaki fark manidar bulunmuştur ( $Q=13,29$ ,  $p < .05$ ). Bu fark, ikinci kesme puanının belirlenmesinde yargıcılar arasında uyum olmadığı anlamına gelmektedir. Bu işlemler geçerlik sınavı yapılırken ölçüt olarak kabul edilecek ikinci madde haritalama çalışmasının yargıcılar arası uyum anlamındaki güvenilirliğine ilişkin kanıt bulmak için yapılmış, bu çalışmanın büyük ölçüde güvenilir olduğu görülmüştür.

Geçerlik kanıtı elde etmek üzere yapılan, iki ayrı madde haritalama işleminin uyumunu araştırmaya yönelik yapılan Wilcoxon İşaretili Sıralar Testi sonuçları Tablo 4, Tablo 5 ve Tablo 6'da verilmiştir.

Tablo 4'te, iki ayrı madde haritalama işlemi sonunda belirlenen 1.kesme puanları arasında manidar bir fark olmadığı görülmektedir ( $z=-1,19$ ,  $p > .05$ ).

**Tablo 4**

*MTK'ya Dayalı İki Ayrı Madde Haritalama İşlemine Göre Belirlenen 1.kesme Puanları Arasındaki Farkın Manidarlığına İlişkin Wilcoxon İşaretli Sıralar Testi Sonuçları*

	<b>n</b>	<b>Sıra ortalaması</b>	<b>Sıra toplamı</b>	<b>z*</b>	<b>p</b>
Negatif Sıra	5	4,2	21	-1,19	0,23
Pozitif Sıra	2	3,5	7		
Eşit	0				

*Not: \*pozitif sıralar temeline dayalı*

Tablo 5'e göre, iki madde haritalama yöntemi ile belirlenen 2.kesme puanları arasında manidar bir fark olmadığı görülmektedir ( $z=-1,27$ ,  $p > .05$ ).

**Tablo 5**

*Üniversite Öğrencilerinin Akademik E-Usulsüzlük Yapma Durumuna İlişkin Aldıkları Puanlar*

	<b>n</b>	<b>Sıra ortalaması</b>	<b>Sıra toplamı</b>	<b>z*</b>	<b>p</b>
Negatif Sıra	6	4,5	27	-1,27	0,21
Pozitif Sıra	2	4,5	9		
Eşit	0				

*Not: \*pozitif sıralar temeline dayalı*

Tablo 6'ya göre de iki ayrı madde haritalama işleminden elde edilen kesme puanları arasında manidar bir fark bulunmamıştır ( $z=-1,27$ ,  $p > .05$ ). Bu farksızlıklar iki ayrı madde haritalama işlemine göre belirlenen 1., 2. ve 3. kesme puanlarının benzer olduğunu göstermektedir. Bu durum MTK'ya dayalı MH yöntemiyle 3 ayrı kesme puanının belirlenmesine ilişkin geçerlik kanıtı olarak kabul edilmiştir.

**Tablo 6**

*MTK'ya Dayalı İki Ayrı Madde Haritalama İşlemine Göre Belirlenen 3.kesme Puanları Arasındaki Farkın Manidarlığına İlişkin Wilcoxon İşaretli Sıralar Testi Sonuçları*

	<b>n</b>	<b>Sıra ortalaması</b>	<b>Sıra toplamı</b>	<b>z*</b>	<b>p</b>
Negatif Sıra	5	43,40	17	-0,14	0,89
Pozitif Sıra	3	6,33	19		
Eşit	0				

*Not: \*negatif sıralar temeline dayalı*

Angoff ve MTK'ya dayalı MH Yöntemlerine ilişkin belirlemeler için güvenilirlik kanıtları elde etmek için yargıcılar arası tutarlılık araştırılmıştır. Angoff yöntemine göre yargıcıların belirledikleri kesme noktaları altında kalan madde sayıları aynı zamanda yargıcıların bireysel olarak belirledikleri ham kesme puanlarına (80 puan üzerinden) karşılık gelmektedir. Yargıcı kararlarına göre 1.kesme puanı 5-45, 2.kesme puanı 8-66 ve 3.kesme puanı da 57-80 arasında değişmektedir.

Angoff yöntemine göre yapılan belirlemelerde 1., 2. ve 3. kesme puanlarının belirlenmesinde yargıcıların verdikleri kararlar arasında manidar bir fark ( $Q_1=88,76$ ,  $Q_2=132,24$ ,  $Q_3=48,00$ ;  $p < .05$ ) vardır. Bu durum yargıcıların her üç kesme puanı belirlenmesi için de uyum göstermedikleri biçiminde yorumlanabilir. Yargıcıların kararlarındaki uyumsuzluk onların tanımlanan adaylara ilişkin beklentilerinin birbirlerinden farklı olmasına ya da maddelerin güçlükleri konusundaki düşüncelerinin farklılığına bağlanabilir.

MTK'ya dayalı madde haritalama yönteminde ise yargıcıların 1.kesme puanının altına yerleştirdiği madde sayılarının 5 ile 7; 2.kesme puanının altına yerleştirdiği madde sayılarının 9

ile 11; 3.kesme puanı altına yerleştirdiği madde sayılarının ise 13 ile 15 arasında değiştiği görülmektedir. Bu yönetime göre yapılan belirlemelerde, 1., 2. ve 3. kesme puanının belirlenmesine ilişkin uyumu araştırmak için yargıcıların verdikleri kararlar arasındaki farklara bakılmış ve bu farkların manidar olmadığı görülmüştür ( $Q_1=8$ ,  $Q_2=8,67$ ,  $Q_3=8$ ;  $p > .05$ ). Bu durumda yargıcıların 1., 2. ve 3. kesme puanlarının belirlenmesinde birbirleri ile uyumlu kararlar verdikleri belirlenmiştir.

Angoff ve MTK'ya dayalı MH yöntemlerine dayalı olarak belirlenen kesme puanlarının karşılaştırılmasına ilişkin bulguda Tablo 7'den yararlanılmıştır

**Tablo 7**

*Angoff ve MTK'ya Dayalı MH Yöntemlerine Göre Belirlenen Kesme Puanları Arasındaki Farklar*

Değişken	Angoff	MTK'ya Dayalı MH	Fark
1.kesme puanına denk gelen madde sayısı	23	37	14
tüm maddelere oranı	%23,28	%46,25	%17,5
2. kesme puanına denk gelen madde sayısı	51	49	2
tüm maddelere oranı	%63,75	%61,25	%2,5
3. kesme puanına denk gelen madde sayısı	77	77	0
tüm maddelere oranı	%96,25	%96,25	%0

Tablo 7'ye göre MTK'ya dayalı MH yöntemine göre belirlenen 1.kesme puanı Angoff yöntemi ile belirlenene göre 14 puan daha yüksektir. Belirlenen 2.kesme puanlarına bakıldığında Angoff yöntemine göre belirlenen kesme puanının MTK'ya dayalı MH yöntemine göre belirlenen kesme puanından 2 puan daha yüksek olduğu görülmektedir. 3.kesme puanı ise her iki yöntemle de aynı ham puana denk gelmektedir.

İki yönetime göre belirlenen kesme puanları arasındaki farkın manidarlığı araştırıldığında, Angoff ve MTK'ya dayalı MH yöntemlerine göre 1.kesme puanının altında kalan maddelerin yüzdeleri arasındaki %17,5'lik farkın anlamlı olduğu bulunmuştur ( $z=2,71$ ;  $p>.05$ ). Bu bulgu ile Angoff ve MTK'ya dayalı Madde Haritalama yöntemlerine göre "öğretim üyesi dışındaki öğretim elemanı kadrolarına başvurmaya hak kazanacak İngilizce açısından minimum yeterlik düzeyindeki aday" a yönelik belirlenen kesme puanları arasında manidar fark olduğu belirlenmiştir. Bu durum yargıcıların OİB ve madde haritası yolu ile uygulama sonuçlarına ilişkin bilgi edinmelerinin 1.kesme puanı için referans alınan adaylara ilişkin beklentilerini yükselttiğini dolayısıyla MTK'ya dayalı MH yöntemiyle Angoff yönteminden daha yüksek kesme puanı belirlediklerini göstermektedir.

Angoff ve MTK'ya dayalı MH yöntemlerine göre belirlenen 2.kesme noktalarının altında kalan maddelerin yüzdeleri arasında manidar bir fark olmadığı bulunmuştur ( $z=0,35$ ,  $p>.05$ ). Bu bulgu, "Doçentlik başvurusu yapmaya hak kazanacak İngilizce açısından minimum yeterlikteki aday" a yönelik Angoff ve MTK'ya dayalı MH yöntemleri ile belirlenen kesme puanlarının benzer olduğu anlamına gelmektedir. Angoff ve MTK'ya dayalı Madde Haritalama yöntemlerine göre yapılan belirlemeler sonucunda 3.kesme puanının altında ve üstünde kalan madde sayıları aynı olduğundan herhangi bir karşılaştırmaya gidilmemiştir. İki yöntemle, "İngilizce okutmanlığı başvurusu yapmaya hak kazanacak aday" a yönelik aynı kesme puanı belirlenmiştir.

## TARTIŞMA ve SONUÇ

Bu araştırmada, Türkiye'de ÖSYM tarafından yılda iki kez uygulanan ve akademik kadrolara atanma aşamasında başvuru koşulu olarak kabul edilen ÜDS'nin 2005 yılına ait Sosyal Bilimler için İngilizce testi üzerinde çalışılmış ve bu test için ÖSYM tarafından ÜDS'nin kullanılmasına yönelik belirlenen üç amaç için Angoff ve MTK'ya dayalı MH yöntemleri ile kesme puanları

belirlenmesi amaçlanmıştır. Araştırmanın amaçları doğrultusunda elde edilen sonuçlar aşağıda özetlenmiştir;

MTK'ya dayalı MH ve Angoff yöntemleri ile belirlenen 1.kesme puanları birbirinden farklı iken, 2. ve 3. kesme puanlarının birbirine yakın olduğu görülmüştür. Genel olarak iki yönteme göre benzer kesme puanları elde edilmiştir. MTK'ya dayalı belirlemelere bir geçerlik kanıtı oluşturmak amacıyla tüm gruptan random olarak bir yanıtlayıcı örnekleme seçilmiş bu gruba ait verilerden elde edilen madde parametrelerine göre maddeler yeniden haritalanmış ve yargıcılarla ikinci bir madde haritalama çalışması yürütülmüştür. Bu iki uygulamada verilen kararların birbirleri ile uyumlu olması, yapılan işlemin çapraz geçerliği için kanıt olarak kabul edilmiştir.

Angoff ve MTK'ya dayalı MH yöntemleri ile yapılan belirlemelerin güvenilirliğine ilişkin bir kanıt elde etmek amacıyla her iki yöntem için üç ayrı kesme puanının belirlenmesinde yargıcılar arası uyum araştırılmıştır. Buna göre MTK'ya dayalı MH yönteminde yargıcılar üç kesme puanı için de birbirleri ile uyumlu kararlar verirken Angoff yönteminde kesme puanlarının üçü için de yargıcı kararları birbirleri ile uyum göstermemektedir. Bu durum, madde haritasının ve sıralanmış madde kitapçığının yargıcılara maddelerin güçlük sıraları konusunda fikir vermesine bağlanabilir. MTK'ya dayalı belirlemelerde yargıcılar gerçek uygulama sonuçlarına dayalı madde güçlükleri konusunda madde haritası ve SMK ile yönlendirilirken Angoff Yönteminde yargıcılar madde güçlükleri konusunda tamamen kendi tahminlerini kullanarak kararlar vermektedirler. Hiç bir dayanak olmadan yapılan bireysel tahminler, yargıcı kararlarının ortak standarttan uzaklaşmasına neden olmuştur. Bu çalışmanın bulgularına benzer biçimde, Skorupski ve Hambleton (2005)'in çalışmasında yargıcıların Angoff yönteminde kendi içlerinde bile tutarsızlıklar yaşadıkları görülmüş, bu durum yargıcıların görevlerini iyi anlayamamalarına bağlanmıştır. Ancak farklı bulgular da söz konusudur. Örneğin Impara ve Plake (1998)'in araştırmasında Angoff yöntemindeki yargıcıların madde performansını tahmin etmesine ilişkin, yargıcılar arası ve uyum olduğu görülmüştür. Söz konusu çalışmada yargıcılara tartışma olanağı sunulmuş olmasının onların birbirleri ile uyumlu kararlar vermelerini sağladığı düşünülebilir.

Angoff ve MTK'ya dayalı Madde Haritalama yöntemlerine göre belirlenen kesme puanları arasındaki farkın manidarlığı araştırıldığında, iki ayrı yönteme göre belirlenen kesme puanlarından sadece 1.kesme puanının altında kalan maddelerin yüzdeleri arasında anlamlı fark olduğu görülmüştür. Bu çalışmada 1. kesme puanı için, aynı yargıcıların yöntem değiştiğinde farklı karar vermeleri onların bu yeterlik düzeyi için tanımlanan sınır aday grubuna yönelik kararlarının net olmadığı, bu belirsizlik sebebiyle gerçek verilere göre belirlenmiş madde güçlüğü sıralarından kolayca etkilendikleri biçiminde yorumlanabilir. Skaggs ve Tessema (2001)'nin çalışması da yargıcılara madde güçlük sıraları konusunda bilgi verilmesinin onların kararını önemli derecede etkilediğini göstermektedir. 2. ve 3.kesme puanının belirlenmesinde iki yöntem arasında uyum olması ise yargıcıların madde güçlüklerine ilişkin yargılarının gerçek veriler ile uyum sağladığı, ya da gerçek verilere ilişkin madde güçlükleri sıralamasını gözetmelerinin kararlarını değiştirmede göstermektedir.

Kesme puanlarının belirlenmesinde farklı yöntemlerin birbirleri ile uyumu araştırmalara konu olmuş bazı araştırmalar (Irwin 2007) farklı yöntemlerle belirlenen kesme puanlarının farklı sonuçlar verdiğini gösterirken bazı araştırmalarda (Tülübaş, 2009) ise farklı yöntemlere göre belirlenen kesme puanlarının benzer olduğu bulgusuna ulaşılmıştır. Buckhendal, Smith, Impara ve Plake (2002)'in çalışmasında MTK'ya dayalı işaretleme yöntemi ile Angoff yöntemi karşılaştırılmış ve Angoff yöntemiyle daha düşük kesme puanları elde edildiği sonucuna ulaşılmıştır. Buna karşın, Wang ve diğerleri (2001)'nin çalışmasının bulgusu, Angoff yöntemi ile belirlenen kesme puanlarının, MTK'ya dayalı MH yöntemi ile elde edilen kesme puanlarından daha yüksek olduğu yönündedir. Literatürde her iki yönde de fark olması, elde edilen bulguların yapılan işlemlere ve teste göre farklılaşabildiğini göstermektedir.

Sonuç olarak, bu araştırma kapsamında ÜDS 2005 Sosyal Bilimler Testi için Angoff ve MTK'ya dayalı MH yöntemlerine göre kesme puanları belirlenmiş, iki yöntemle dayalı belirlemelerin birbirleri ile genel olarak uyumlu olduğu ancak MTK'ya dayalı MH Yöntemi ile yapılan belirlemelerin daha güvenilir olduğu sonucuna ulaşılmıştır. Ayrıca, yargıcıların iş yükünü azaltma ve onlara ortak standardı yakalama konusunda yol gösterme gibi özellikleri sebebiyle MTK'ya dayalı MH Yöntemi'nin Angoff yöntemine göre daha kullanışlı olduğu görülmüştür. Çalışmada, MTK'ya dayalı madde haritalama çalışması için bir geçerlik kanıtı elde edilmiş, böylece kesme puanı belirlemeye ilişkin model önermenin yanı sıra belirlenen kesme puanlarının geçerlik ve güvenilirliğine ilişkin kanıt elde etmeye yönelik yollar da açıklanmıştır.

Bu çalışmanın bulguları doğrultusunda MTK'ya dayalı MH Yöntemi'nin daha güvenilir olduğu dikkate alınarak, kesme puanları belirlenirken yargısal yaklaşım ile istatistiksel yaklaşımı birleştiren MTK'ya dayalı MH yönteminin kullanılması önerilebilir. Ancak MTK'nın avantajlarından yararlanabilmek için MTK varsayımlarının karşılanması ve maddelerin MTK'ya uyum sağlaması gerektiğinden testlerin geliştirilmesi aşamasında da MTK'dan yararlanılması önerilir. Bu yöntemle göre sürdürülen standart belirleme çalışmalarında özet harita kullanılmasının sonuçları değiştirmedeği görülmüştür. Bu durumda, yargıcıların iş yükünü azaltmak ve zamandan tasarruf etmek için tüm maddelerin yer aldığı yüklü ve karmaşık haritalar yerine temsilci maddelerden oluşan özet madde haritalarının kullanılması önerilebilir. Hem MTK'ya dayalı MH hem de Angoff Yöntemi ile kesme puanı belirlemesi yapılırken, bu çalışmada olduğu gibi kesme puanlarının belirlenmesinde esas alınacak adayların, yargıcı grubundan bağımsız bir uzman grubu tarafından tanımlanması yargıcılar için bu adayların ölçülen özellik açısından durumlarını daha somut ve anlaşılır kılacaktır.

Ayrıca, mutlak değerlendirmeyi esas alan testlerde kesme puanları arasında yer alan düzeylerin tanımlanması bu kesme puanları açısından sayısal bilginin yanında nitel dönüt de verilmesini sağlayacağından kesme puanları arasında kalan yeterli düzeylerinin belirlenmesi ve testi alanlar ile test sonuçlarını kullananların bilgilendirilmesi önerilebilir. Özellikle ÜDS gibi periyodik olarak uygulanan sınavların yeterli düzeylerinin tanımlanması, bir sonraki periyotta aynı sınavı alacak olanlar için yol gösterici olacaktır.

### **Araştırmanın Sınırlılıkları**

Bu çalışmada ÖSYM tarafından maddeleri ve yanıtlayıcı verileri paylaşılan geçmişte yapılan bir sınavın, 2005 yılına ait ÜDS Sosyal Bilimler Testi'nin verileri kullanılmıştır. MTK'ya ilişkin analizlerin, modele uyum gösteren maddeler ile gerçekleştirilmesi gerekmesine rağmen, hazır bir testle çalışılması sebebiyle ve bu testin bütünlüğünü bozmamak adına MTK'ya uyum göstermeyen maddeler de analizlere dâhil edilmiştir. Ayrıca bu çalışmada yargıcıların iş yükünün fazla olması ve araştırmacının olanakları gözetilerek yargıcı sayısı yedi ile sınırlı tutulmuştur.

### **Destek ve Teşekkür**

Bu makale "Angoff ve Madde Tepki Kuramına Dayalı Madde Haritalama Yöntemleri İle ÜDS (Üniversiteler Arası Kurul Yabancı Dil Sınavı) için Yeterlik Standartlarının Belirlenmesi" isimli doktora tezinden üretilmiştir.

### **Araştırmacıların Katkı Oranı**

Araştırmanın birinci yazarı tez yazarı olup ikinci yazarı ise bu tez sürecine danışmanlık yapmıştır.

### **Çatışma Beyanı**

Araştırmanın yazarları olarak herhangi bir çıkar/çatışma beyanımız olmadığını ifade ederiz.

## Yayın Etiği Beyanı

Bu araştırmanın planlanmasından, uygulanmasına, verilerin toplanmasından verilerin analizine kadar olan tüm süreçte “Yükseköğretim Kurumları Bilimsel Araştırma ve Yayın Etiği Yönergesi” kapsamında uyulması belirtilen tüm kurallara uyulmuştur. Yönergenin ikinci bölümü olan “Bilimsel Araştırma ve Yayın Etiğine Aykırı Eylemler” başlığı altında belirtilen eylemlerden hiçbiri gerçekleştirilmemiştir.

Bu çalışmanın yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulmuş; toplanan veriler üzerinde herhangi bir tahrifat yapılmamış ve bu çalışma herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiştir.

## KAYNAKÇA

- Akhun, İlhan. (1982). İki yüzde arasındaki farkın manidarlığının test edilmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 15(1), 240-259.
- American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (2002). *Standards for educational and psychological testing*. Washington: AERA.
- Bachman, Lyle F. (1990). *Fundamental considerations in language testing* (2. Basım). Oxford: Oxford University Press.
- Buckhendal, W. C., Smith W. R., Impara C. J., & Plake S. B. (2002). A comprasion of Angoff and bookmark standard setting methods. *Journal of Educational Measurement*, 39(3), 253-263.
- Camilli, G., Cizek, G. J., & Lugg, C. A. (2001). Psychometric theory and the validation of performance standards. In G. J. Cizek (ed.) *Setting Performance Standards: Concepts, Methods and Perspectives* (445-475). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cizek, Gregory J. (ed.). (2001). *Setting performance standards: Concepts, methods and perspectives*. NJ: Lawrence Erlbaum Associates.
- Cizek, G., & Bunch M. (2004). Setting performance standards: Contemporary methods. *Educational Measurement: Issues and Practice*, 23(4), 31-50.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. NY: CBS College Publishing Company.
- DeMauro, G. E. (1995, May). *Construct validation of minimum competence a bookmark standard setting procedure*. Paper presented at the Annual Meeting of the National Council on Measurement in Education. Francisco, CA.
- Dickinson, E. R., Thacker, A. A., Moody, R. K., & Hoffman, R.G. (2006). *Item content and difficulty mapping by form and item type for the 2006 Kentucky Core Content Tests*. Louisville: Human Resources Research Organization.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Gwet, Kilem L. (2001). *Statistical tables for inter-rater agreement*. Gaithersburg: StatAxis Publishing.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Boston, Kluwer - Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1996). *Fundamentals of item response theory*. California: SAGE Pub.
- Impara, J. C., & Plake, B. S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34(4), 353-366.
- Impara, C. J., & 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.
- Irwin, Patrick M. (2007) *An alternative examinee-centered standard setting strategy* (Yayınlanmamış doktora tezi). USA: University of Nebraska.
- Jaeger, Richard M. (1989) Certification of student competence. In R.L. Linn (Ed.) *Educational Measurement* (3rd ed.), Newyork.
- Karantonis, A., & Sireci, S.G. (2006, Mar). The bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice*, 25(1), 4-12.



- Lissitz, R. W., & Kroopnick, M. H. (2007). *An adaptive procedure for standard setting and a comparison with traditional approaches*. Paper presented at the NCME annual meeting in Chicago, IL.
- Moy, R. H. (March, 1981). *Proficiency standards and cut scores for language proficiency tests*. Paper presented at the Annual Meeting of the TESOL, Detroit.
- Ryan, J. M. (2003). *An analysis of item mapping and test reporting strategies*. [http://www.seirtec.org/downloads/Non\\_REL/Assessment/Analysis\\_of\\_State\\_Test\\_Score\\_Reporting\\_Strategies.pdf](http://www.seirtec.org/downloads/Non_REL/Assessment/Analysis_of_State_Test_Score_Reporting_Strategies.pdf)
- Schulz, Matthew E. (2006). Commentary: A response to reekase's conceptual framework and examples for evaluating standard setting methods. *Measurement: Issues and Practice*, 25(3), 4-13.
- Seçil, S. Ö. (2009). *Matematik öğrenimindeki bilişsel süreçlerin madde tepki kuramıyla incelenmesi*. (Yayınlanmamış doktora tezi). Orta Doğu Teknik Üniversitesi, Ankara.
- Shen, L. (2001, April). *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.
- Skaggs, G., & Tessema, A. (2001, April). *Item disordinality with the bookmark standard setting procedure*. Paper presented at the Annual Meeting of the National Council on Measurement in Education. Seattle, WA.
- Skorupski, W. P., & Hambleton, R. K. (2005). What are panelists thinking when they participate in standard-setting studies? *Applied Measurement in Education*, 18(3), 233-256.
- Tülübaş, G. (2009). *Psikolojik testlerde Angoff ve sınır grup yöntemleri ile kesme puanlarının belirlenmesi*. Yayınlanmamış yüksek lisans tezi, Hacettepe Üniversitesi, Ankara.
- Wang, N., Wiser, R. F., & Newman, L. S. (2001, April). *Use of the Rasch IRT Model in Standard Setting: An Item Mapping Method*. Version of a paper presented at the Annual Meeting of the National Council on Measurement in Education, Seattle, WA.
- ÖSYM. (2020). *Ölçme, Seçme ve Yerleştirme Merkezi*. <http://www.osym.gov.tr>
- Zwick, R., Şentürk, D., & Wang, J. (2001). An investigation of alternative methods for item mapping in the national assessment of educational progress. *Educational Measurement: Issues and Practice*, 20(2), 15-25.

## EXTENDED ABSTRACT

### Introduction

The concept of standard setting is used in the literature to separate the scores of a test to certain levels with the specified cut-off points based on the probability of correct answers of the test items (Irwin, 2007; Schulz, 2006; Shen, 2001; Zwick, Şentürk & Wang, 2001). In this study, it was aimed to determine new cut-off scores, with the Item Mapping based on IRT and Angoff methods for the ÜDS Social Sciences English Test. In Angoff Method which is the subject of this research; Items that are considered to be the minimum answers acceptable are given 1 point and others are given 0 points. The sum of the item scores constitutes the cutoff score (Cizek, 2001).

In standard setting studies, there are also statistical approaches other than those that take only judge decisions into account. The other method in this study, Item Response Theory based Item Mapping, is a standard setting method based on statistical approach. According to IRT, the probability of answering the item correctly is a function of the latent trait, that is, ability. Accordingly, the traits of the respondents and the properties of the items are scaled on the same scale in the theory (Crocker and Algina, 1986). In this way, the probability of a student at a certain level of ability to answer an item that has not yet answered correctly is estimated through the "item characteristic curve", which gives the probability of answering that item correctly for each ability level. In other words, based on the performance in a number of items, the performance in another set of items can be predicted. In IRT based Item Mapping method, since it is possible to make estimates about the items that are not presented to the judges, the judges are presented with the item sample, not the entire items. Thus, estimation can be made with fewer items and saves time and effort. In addition, since it can be predicted who will get which score, the judges will know the passing rates corresponding to the cut-off score they set (Hambleton, Swaminathan and Rogers, 1996; Wang and others, 2001; Ryan, 2003).

### Method

For the study, the archive data of the Social Sciences English test for the year 2005 of ÜDS, which was conducted twice a year by ÖSYM between 2000 and 2012 and whose scores were accepted as a prerequisite during the appointment to academic staff, were used. These data belong to 3731 candidates. In the study, cut-off points announced by ÖSYM were tried to be determined again by Angoff and IRT based Item mapping methods. Seven judges were employed in the execution of both methods. The language competencies of the candidates who are at the limit of their proficiency levels regarding the 3 cut-off points to be determined were defined in line with the opinions of 15 experts, based on the European language portfolio framework.

Within the scope of the general purpose of the study, cut-off scores of the IRT based Item Mapping and Angoff methods were determined for the " assignment to faculty members other than assistant professors, associate professors and professors ", " applying for Associate Professorship " and " applying for English lecturer " proficiency levels in the interpretation of the ÜDS Social Sciences English test scores. While setting the standard with the Angoff method, each items were asked to the judges as "Can the candidate at the border of the cut-off score answer this item?" In the IRT based Item Mapping study, the judges were presented with an item sampling instead of all items, with the help of item maps for each item in the sample, They were asked to answer the same question for each item again "Can the candidate at the border of the cut-off scorer answer this item?"

As a validity proof, it was investigated whether there is a difference between the cut-off scores determined by the IRT based Item mapping for a sample of respondents chosen randomly from the whole group and for the whole group. As a reliability proof, when the Angoff and IRT based Item Mapping methods were used in determining the cut-off points, the coefficients of agreement between judges were calculated. In addition, the significance of the difference between the cut-off points determined by the Angoff and IRT based Item Mapping Methods was also investigated.

## Findings, Discussion and Results

It has been observed that the 1st cut scores determined by IRT based Item Mapping and Angoff methods are different from each other, but the 2nd and 3rd cut points are close to each other. Based on the data of the whole group and the random sample, it was observed that the decisions made by the judges were consistent with each other. This situation was accepted as evidence for the cross validity of the procedure. As proof of reliability, when the consistency between judges in determining three separate cut-off points for both methods was examined, it was found that the judges made concerted decisions for the three cut-off points in the IRT based Item mapping method. However, in the Angoff method, the judge decisions for all three cut-off scores do not comply with each other. It has been observed that there is a significant difference between the percentages of the items below the first cut score from the cut points determined according to the Angoff method and Item Mapping method based on IRT.