

Angoff, Evet/Hayır ve Sınır Grup Yöntemlerine Göre Kesme Puanlarının Karşılaştırılması

Ali Kılıç^a ve Sevilay Kilmen^b

Öz

Bu araştırmada test merkezli standart belirleme yöntemlerinden Angoff ve evet/hayır yöntemleri ile öğrenci merkezli standart belirleme yöntemlerinden sınır grup yöntemi ile elde edilen kesme puanlarının karşılaştırılması amaçlanmıştır. Araştırma kapsamında öğrenci ve yargıcı grubu olmak üzere iki farklı çalışma grubundan veriler elde edilmiştir. Öğrenci grubunu Uşak ili Anadolu liselerinde 10. sınıfta öğrenim gören 1.057 öğrenci, yargıcı grubunu ise öğrenci grubunun dersine giren 22 matematik ve 16 Türk dili ve edebiyatı öğretmeni oluşturmaktadır. Araştırma sonunda tüm yöntemler arasında başarılı kabul edilen öğrenci oranları açısından anlamlı farklar olduğu saptanmıştır.

Makale Hakkında

Gönderim tarihi: 15.01.2019

Düzeltilme tarihi: 04.05.2019

Kabul tarihi: 22.09.2019

Elektronik Yayın Tarihi: 15.12.2019

Anahtar kelimeler: Angoff, evet/hayır, sınır grup, kesme puanı

Giriş

Bireylerin sınava tabi tutulmasının amaçlarından biri de bireylerin test sonuçları hakkında karar verme veya bireyleri sınıflandırma işlemi yapmaktır. Bu işlemler için kullanılacak ölçüt veya kesme puanının belirlenmesi standart belirleme çalışmalarını elde edilmektedir. Standart belirleme, değerlendirmeyi daha sistematik, tekrarlanabilir, objektif ve savunulabilir duruma getirmektedir (Cizek ve Bunch, 2007). Test sonuçları hakkında verilen kararların geçerliğini artırmada standart belirleme süreci önemli bir rol oynamaktadır. Standartları belirli olmayan test sonuçlarının öğrenciler, eğitimciler ve veliler için yorumlanması oldukça zordur. Rotherham (2006)'a göre bu işlem ölçeksiz bir haritayı okumakla eşdeğerdir.

Uygulanan bir eğitimin sonunda bireyler sertifika almakta veya bir programdan mezun olabilmektedir. Ancak bireylerin sertifika alabilmesi veya mezun olabilmesi için belirli koşulları yerine getirmesi gerekmektedir. Bu koşulların yerine getirilip getirilmediği, bireylerin gösterdiği performansların ölçümü ile değerlendirilmesi ile sağlanmaktadır. Değerlendirme yapılırken, bireylerin başarılı olabilmesi için "ne kadar yeterli?" sorusu sorulmaktadır. "Ne kadar yeterli?" sorusunun cevabı aslında standart veya ölçütün bir karşılığıdır (Livingston ve Zieky, 1982).

Standart belirleme süreci zor bir süreç olmasına rağmen dolaylı olarak öğrenci performansını arttıran bir süreçtir. Açıkça belirlenmiş bir standart veya ölçütün varlığı öğrencilerin başarılı olmak için gerekli olan performans düzeyine ulaşmasını

^a Milli Eğitim Bakanlığı, alikilic@windowstlive.com, ORCID: 0000-0001-8782-9097.

^b Sorumlu yazar. Bolu Abant İzzet Baysal Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, kaplansevilay@yahoo.com, ORCID: 0000-0002-5432-7338.

kolaylaştırmaktadır. Öğrenciler ve öğretmenler, standartların belirli olduğu durumlarda gerekli performansa ulaşabilmek için daha çok ilgi ve çaba gösterebilmektedir (Cizek ve Bunch, 2007).

Standart belirleme süreci, öğrenci performansının değerlendirilmesi dışında farklı amaçlar için de kullanılabilir. Bu amaçlarından biri de okullardaki eğitim kalitesinin artırılmasına yöneliktir. Ancak standart belirleme eğitimin geliştirilmesi için tek başına kullanılabilecek bir süreç değildir (Livingston ve Zieky, 1982). Örneğin, aynı okul türünde veya aynı programı uygulayan okullarda standart belirleme çalışmalarında oldukça farklı kesme puanları elde edilebilmektedir. Bu bulgular, bazı okullar, program türleri veya bölgelerin diğerlerine nazaran daha fazla iyileştirmeye ihtiyaç duyduğunu göstermektedir. Standart belirleme çalışmaları tamamlandıktan sonra iyileştirmeye yönelik politikaların belirlenmesi ve birtakım tedbirlerin alınması gerekmektedir. Bu nedenle standart belirleme, durum tespitinden çok, eğitim kalitesinin artmasına yönelik yöntem belirleme amaçlı kullanılmalıdır (Tülübaş, 2009).

Standart Belirleme Aşamaları

Standart belirleme, belirli bir plan ve aşamaları içeren süreçler bütünüdür. Standart belirleme konusunda birçok yöntem bulunmaktadır. Bu yöntemlerin uygulanması için birbirinden farklı süreçler bulunmasına rağmen, belirli süreçler ortak kullanılmaktadır. Literatürde 5 farklı basamaktan 12 farklı basamağa kadar standart belirleme çalışmalarında ortak kullanılan standart belirleme basamağı bulunmaktadır (Wyse, 2009). Norcini (2003)'ye göre ortak olan basamaklar aşağıdaki gibidir.

1. Kullanılacak standart tipinin belirlenmesi: Kullanım amacına göre mutlak ve bağıl olmak üzere iki tip ölçüt veya standarttan bahsetmek mümkündür. Mutlak ölçüt, önceden tanımlanmış, gruptan bağımsız bir eşik değeri iken; bağıl ölçüt grubun test puanları ile elde edilen norm dayanaklı bir ölçüttür (Atılğan, Yurdakul ve Öğretmen, 2012). Mutlak ölçüt kullanımında test maddelerinin belirli bir yüzdesi veya oranı ölçütü oluşturmaktadır. Örneğin 100 sorudan 60'ını doğru yanıtlayan (%60) başarılı kabul edilmektedir. Ancak bağıl ölçüt kullanımında gruptaki öğrencilerin yüzdesi veya oranı ölçütü oluşturmaktadır. Örneğin 100 öğrenci arasından ilk 20 başarılı kabul edilmektedir (Norcini, 2003). Bu bağlamda kullanılacak ölçüt tipi veya standart tipi; kesme puanını doğrudan olarak etkilemektedir. Bu nedenle standart belirleme süreçlerinin başlangıç aşamasını kullanılacak standart tipinin belirlenmesi oluşturmaktadır.

2. Standart belirleme yönteminin seçilmesi: Her yöntemin diğerlerine göre avantajları olsa da birtakım dezavantajları da bulunmaktadır. Standart belirleme yöntemi seçilirken zaman, madde tipi, yöntemin uygulanması için gereken şartlar ve yargıcıların grup hakkındaki ön bilgileri bir bütün olarak değerlendirilmelidir (Hambleton ve Pitoniak, 2006). Berk (1984)'e göre süreçteki basamakların sistematik olması ve kullanılan yöntemin yargıcılar tarafından kolay anlaşılabilir olması yöntem seçimi için en önemli kriterlerdendir.

3. Yargıcıların seçilmesi: Standart belirleme çalışmalarında yargıcılar genelde üç kritere göre seçilmektedir. Bu kriterlerden birincisi, yargıcıların öğrenci grubu hakkında bilgisidir. Özellikle öğrenci merkezli yöntemlerde yargıcıların, öğrencilerin yeteneklerini iyi tanınması gerekmektedir. Ayrıca, test sonucu başarılı sayılabilecek bir öğrencinin sahip olması gereken beceri düzeyini de yargıcılar başarılı

bir şekilde kestirebilmelidir (Livingston ve Zieky, 1982). İkinci kriter ise çalışmalara katılacak yargıcı sayısıdır. Sınıf gibi küçük gruplar için bir öğretmenin yargıcı olarak seçilmesi yeterli olmaktadır. Ancak, daha büyük gruplar için güvenilir sonuçlar elde edilmek isteniliyorsa bir yargıcı tarafından değerlendirme yapılması yeterli olmayacaktır. Bu gibi durumlarda yargıcı sayısının 10 ile 15 kişi arasında olması önerilmektedir (Hurtz ve Hertz, 1999; Taşdelen, 2009). Üçüncü kriter, yargıcıların demografik yapısı ile ilgidir. Testin amacına bağlı olarak yargıcıların cinsiyet, yaş ve coğrafi bölgeler gibi farklı gruplardan dengeli bir dağılımla seçilmesi; yargıcıların kendi aralarında yapacakları tartışmalar açısından standart belirleme sürecine önemli katkılar sağlamaktadır (Raymond ve Reid, 2001).

4. Standart belirleme toplantılarının düzenlenmesi: Toplantıların amacı yargıcıların kullanılacak yöntem ve süreç hakkında bilgilendirilmesidir. Bilgilendirmeler sayesinde çalışma sonucu ulaşılan kesme puanının güvenilirliği artacaktır. Toplantılarda çalışmanın amacı, kullanılacak test maddeleri, cevap anahtarı, standart belirleme sürecinin genel hatları hakkında yargıcılara bilgi verilmektedir. Bu toplantılarda gündeme alınması gereken diğer bir önemli konu ise performans düzeyleri ve minimum yeterlik düzeyi hakkında yargıcıların bilgilendirilmesidir (Wyse, 2009). Ayrıca; yargıcıların toplantılar boyunca geribildirim alması ve kendi fikirlerini beyan etmesi, sürecin daha sağlıklı ilerlemesine olanak sağlamaktadır (Norcini, 2003).

5. Kesme puanının hesaplanması: Kesme puanının hesaplanması, yargıcılardan verilerin toplanması ile başlayan süreçtir. Genelde minimum yeterlik düzeyindeki öğrencilerin test puanlarından yararlanarak kesme puanı elde edilmektedir. Kullanılan yöntemle ilgili olarak farklı istatistikî değerler hesaplanırsa da yaygın olarak bu puanların aritmetik ortalaması veya medyan değeri kesme puanı olarak kullanılmaktadır. Özellikle öğrenci sayısının yeterli olmadığı durumlarda medyanın uç değerlerden daha az etkilenmesi nedeniyle medyan değeri daha çok tercih edilmektedir (Livingston ve Zieky, 1982; Zieky, Perie ve Livingston, 2006).

6. Sürecin değerlendirilmesi: Bu aşamada standart belirleme süreci 3 farklı yönden incelenmektedir. Bunlardan ilki sonuçların güvenilirliğinin sorgulanmasıdır. İkincisi başarı oranları arasında kıyaslamalar yapılarak geçerliğinin saptanmasıdır. Üçüncüsü ise elde edilen bilgiler ile öğrencilerin gelecekte gösterecekleri performansların kıyaslanması aşamasıdır.

Çalışma Hakkında

Bireyleri bir programa yönlendirirken veya uygulanan programın hedeflerine ne derece yaklaşıldığının belirlenmesi için çeşitli kararlara ihtiyaç duyulmaktadır. Karar verme süreçlerinde ise kesme puanının büyük bir öneme sahip olduğu birçok çalışmada saptanmıştır (Cizek, 2001; Hambleton, 2001; Livingston ve Zieky, 1982). Bu bağlamda kesme puanının hangi yöntemde daha düşük veya hangi yöntemde daha yüksek olduğunun belirlenmesi, öğrencilerin uygulanan yöntemle ilgili olarak başarılı olma durumlarının incelenmesi ve kesme puanlarının kullanılan yöntemle ilgili olarak nasıl değiştiğinin araştırılması; alanyazına ve uygulayıcılara önemli katkılar sağlayacaktır. Bu araştırmada Angoff, evet/hayır ve sınır grup yöntemlerine göre elde edilen kesme puanlarının farklılaşıp farklılaşmadığı araştırılmaktadır. Bu amaçla aşağıdaki sorulara yanıt aranmıştır:

- 1) Angoff, evet/hayır ve sınır grup yöntemlerine göre kestirilen kesme puanları, bu kesme puanlarına göre belirlenen başarılı-başarısız öğrenci sayıları ve yüzdeleri nedir?
- 2) Angoff, evet/hayır ve sınır grup yöntemlerine göre başarılı kabul edilen öğrenci oranları arasında anlamlı bir fark var mıdır?

Standart belirleme yöntemleri test ve öğrenci merkezli olmak üzere ikiye ayrılmaktadır (Jaeger, 1989). Standart belirlemeye yönelik literatür incelendiğinde test merkezli yöntemlerin karşılaştırılmasına yönelik birçok çalışmaya rastlanılmıştır (Cross, Impara, Frary ve Jaeger, 1984; Demir, 2014; Doğan ve Gündeğer, 2014; Hsieh, 2013; Kara ve Kelecioğlu, 2015; Taşdemir, 2013). Ancak, Türkiye’de test ve öğrenci merkezli yaklaşımların her ikisinin de bir arada kullanıldığı araştırmalar sınırlıdır. Bu bağlamda, öğrenci merkezli sınır grup yöntemi ile test merkezli Angoff ve evet/hayır yöntemlerinden elde edilen kesme puanlarının karşılaştırılmasının literatüre ve uygulayıcılara önemli katkılar sağlayacağı düşünülmektedir.

Kavramsal Çerçeve

Angoff yöntemi. Angoff (1971) tarafından geliştirilen yöntem, sınır grupta yer alan bireylerin bir maddeyi doğru cevaplama olasılığına dayanmaktadır (Shin, 2014). Yöntem uygulanış bakımından oldukça kullanışlı bir yöntemdir. Yargıcılar, test maddelerini inceleyerek sınır gruptaki bireylerin maddeleri doğru yanıtlama olasılıklarını yüzdelik olarak belirler, her yargıcının belirlediği bu olasılıklar ayrı ayrı toplanır ve madde sayısına bölünür. Bu elde edilen değer tek bir yargıcının elde ettiği minimum geçme puanıdır (MGP). Tüm yargıcıların elde ettiği minimum geçme puanlarının aritmetik ortalaması ise testin MGP’ni oluşturur.

Angoff yönteminde farklı yaklaşımlarla MGP’na ulaşılabilmektedir. Yargıcılara, “Sınır grupta yer alan 100 öğrenciden kaçının o maddeyi doğru cevaplayabileceği?” sorulabildiği gibi “Sınır grupta yer alan öğrencilerin yüzde kaç olasılıkla o maddeyi doğru cevaplayabileceği?” sorusu da sorulabilmektedir (Demir, 2014). Her iki şekilde de madde için 0 ile 100 arasında bir değer hesaplanır. Maddeye ilişkin bu değer madde güçlük indeksi ile benzer özellik göstermektedir. 0 değerine yaklaşan bir madde için zor, 100 değerine yaklaşan bir madde için ise kolay bir madde değerlendirilmesi yapılabilir.

Angoff yöntemi ile kesme puanının hesaplanması için diğer bir yaklaşımda ise yargıcılara, “Sınır grupta yer aldığı düşünülen herhangi bir öğrenci tarafından yüzde kaç olasılıkla o maddenin doğru cevaplanabileceği?” sorulmaktadır (Demir, 2014). Ancak, bu yaklaşımda yargıcıların sınır grupta olduğunu düşündüğü öğrenciyi doğru tahmin edebilmesi büyük önem taşımaktadır. Bu yaklaşım tek bir kişi üzerinden değerlendirme yapılması nedeniyle diğerlerine nazaran daha kolay uygulanmaktadır. Fakat bu yaklaşım, sınır gruptaki kişinin gerçekte sınır grupta yer almaması durumunda MGP’nin yanlış hesaplanmasına yol açabilmektedir. 5 maddelik ve 100 puan üzerinden değerlendirilen bir testin, 5 yargıcı tarafından Angoff yöntemine göre uygulanması ve MGP’nin hesaplanması Tablo 1’de örnek olarak verilmiştir.

Tablo 1. Angoff yöntemi uygulama örneği

Yargıcılar	Madde-1	Madde-2	Madde-3	Madde-4	Madde-5	MGP
Yargıcı-1	70	80	90	85	75	80
Yargıcı-2	60	70	55	50	45	56
Yargıcı-3	65	75	70	60	50	64
Yargıcı-4	60	60	50	55	45	54
Yargıcı-5	30	35	45	50	20	36
Minimum Geçme Puanı						58

Tablo 1 incelendiğinde, Yargıcı-1, maddelerin sınır gruptaki öğrenciler tarafından doğru yanıtlanma olasılıklarını sırasıyla %70, %80, %90, %85 ve %75 olarak belirlemiştir. Tüm bu değerler toplandığında (70+80+90+85+75) 400 değeri elde edilmektedir. Bu değer madde sayısına bölümü ile (400/5) 80 değerine ulaşılır. Bu 80 değeri, Yargıcı-1'den elde edilen MGP'dir. Diğer yargıcılar için aynı işlem tekrarlandığında MGP'leri sırasıyla Yargıcı-2 için 56, Yargıcı-3 için 64, Yargıcı-4 için 54 ve Yargıcı-5 için 36'dır. Tüm yargıcılardan elde edilen MGP değerlerinin aritmetik ortalaması ise 58 puandır. Bu puan testin MGP'dir. 100 puan üzerinden değerlendirilen bir test için Angoff yöntemi eşitlik olarak ifade edildiğinde;

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

Bu denklemde yer alan terimler; τ_{cj} : j yargıcısı tarafından belirlenen kesme puanı, p_{ij} : sınır grupta yer alan bir öğrencinin j yargıcısı tarafından i maddesini doğru cevaplama yüzdesi, n: testteki madde sayısıdır (Çetin ve Gelbal, 2010). Angoff ve test merkezli yöntemlerin bazıları (örn., Ebel ve Nedelsky yöntemi) uygulanırken, yargıcılara maddelere ilişkin doğru yanıtların ne zaman verileceği bir ikilem oluşturmaktadır. Bu durum şu şekilde açıklanmaktadır. Maddelerin doğru yanıtları yargıcılara değerlendirmeden önce verildiği zaman maddeleri yargıcılar olduğundan daha kolay olarak algılayabilmektedir. Bu uygulamanın sonucu olarak MGP daha yüksek hesaplanmaktadır. Yargıcılara, değerlendirmeden önce doğru yanıtların verilmediği uygulamada ise yargıcılar tarafından maddenin yanlış yanıtlanma olasılığı ortaya çıkmaktadır. Bu uygulamada ise madde kusurlu hale gelebilmektedir. Bu ikilemin çözümü, maddelerin doğru yanıtları verilmeden testin gözden geçirilmesi ve daha sonra doğru yanıtlar verilerek nihai değerlendirmenin yapılmasıdır (Livingston ve Zieky, 1982).

Angoff yöntemi kullanışlı bir yöntem olmasına rağmen, yöneme yönelik birkaç temel eleştiri yapılmıştır. Bu eleştirilerden biri, yöntemde yargıcıların sınır gruptaki öğrencileri baz alarak değerlendirme yapmasıdır. Sınır gruptaki öğrenciler, varsayım olarak geçme sınırının hemen üstündeki veya başarılı ve başarısız performansın tam ortasında yer alan öğrencilerdir. Ancak, bu öğrencileri belirlemek yargıcılar için oldukça güç bir işlemdir (Hsieh, 2013; Plake, 1998). Bu nedenle yöntem uygulanırken yargıcılara sınır grubun net bir şekilde tanımlanması gerekmektedir.

Angoff yöntemine yapılan eleştirilerden bir diğeri de test sonuçları elde edilmeden yargıcılar tarafından maddeler hakkında yapılan değerlendirmelerin güvenilir sonuçlar vermediğine ilişkindir (Clauser, Mee ve Margolis, 2013). Busch ve Jaeger (1990)'a göre yargıcılar tarafından kestirilen maddenin doğru yanıtlanma olasılığı ile test sonuçlarından elde edilen madde güçlük indeksi arasında düşük veya orta düzeyde bir ilişki bulunmaktadır. Bu durum, yargıcılar tarafından yapılan değerlendirmelerin sorgulanmasına neden olmaktadır.

Angoff yönteminin birbirinden farklı modifikasyonları mevcuttur. Bu modifikasyonların uygulanışları birbirinden farklı olmakla beraber, Brandon (2004)'a göre tüm yöntemler için ortak 5 aşama bulunmaktadır. Bu aşamaların ilk dört tanesi yargıcıların seçimi ve eğitimi; performans düzeylerinin tanımı ve her madde için doğru yanıtlanma olasılığının hesaplanmasıdır. Beşinci aşama ise test sonuçları ile yargıcılardan elde edilen MGP'nin karşılaştırılmasıdır.

Evet/hayır (yes/no) yöntemi. Angoff yönteminin bir modifikasyonu olan evet/hayır yöntemi, Impara ve Plake tarafından geliştirilmiştir (Cizek ve Bunch, 2007). Bu yöntem, sınır gruptaki öğrencilerin testte yer alan bir maddeyi doğru cevaplayıp cevaplayamayacağı varsayımına dayanır. Yargıcılar, sınır gruptaki öğrenciler tarafından doğru cevaplanacağı düşünülen her madde için 1, yanlış cevaplanacağı düşünülen her madde için 0 puanını kullanır. Maddelere verilen 1 değerleri her yargıcı için ayrı ayrı toplanır, toplam madde sayısına bölünür ve daha sonra 100 ile çarpılır. Bu elde edilen değer tek bir yargıcının elde ettiği MGP'dir. Tüm yargıcıların elde ettiği MGP'lerinin aritmetik ortalaması ise testin MGP'nı oluşturur. 5 maddelik ve 100 puan üzerinden değerlendirilen bir testin, 5 yargıcı tarafından evet/hayır yöntemine göre uygulanması, aşağıdaki tabloda örnek olarak verilmiştir.

Tablo 2. Evet/hayır yöntemi uygulama örneği.

Yargıcılar	Madde-1	Madde-2	Madde-3	Madde-4	Madde-5	MGP
Yargıcı-1	1	0	1	0	1	60
Yargıcı-2	0	0	1	0	1	40
Yargıcı-3	1	1	1	1	1	100
Yargıcı-4	1	1	0	1	1	80
Yargıcı-5	1	0	0	0	1	40
Minimum Geçme Puanı						64

Tablo 2 incelendiğinde, 5 maddeden oluşan bir test için yargıcılar, maddelerin sınır gruptaki bir öğrenci tarafından doğru yanıtlanıp yanıtlanmayacağını tahmin etmektedirler. Yargıcı-1, maddeler için sırasıyla 1,0,1,0,1 puanlarını kullanmıştır. Tüm puanlar toplandığında (1+0+1+0+1) 3 değeri elde edilmektedir. Bu değer toplam madde sayısına bölümü (3/5) 0,6'dır. Bu sayının 100 ile çarpımından (0,6x100) 60 değerine ulaşılmaktadır. Bu değer Yargıcı-1 için MGP'dir. Diğer yargıcılar için aynı

işlem tekrarlandığında MGP'ları sırasıyla Yargıcı-2 için 40, Yargıcı-3 için 100, Yargıcı-4 için 80 ve Yargıcı-5 için 40'dır. Tüm yargıcılardan elde edilen MGP değerlerinin aritmetik ortalaması ise 64 puandır. Bu puan testin MGP'dir.

Evet/hayır yöntemi, diğer Angoff modifikasyonlarına göre yargıcılar tarafından daha kolay uygulanabilmektedir. Bu durumun nedeni, diğer yöntemlerde yargıcılar sınır grupta yer alan tüm öğrenciler üzerinden maddenin doğru cevaplama olasılığını belirlemekte iken; evet/hayır yönteminde yargıcılar sınır grupta yer alan tek bir öğrenci üzerinden de bu olasılığı belirleyebilmektedir. Tek bir öğrenci üzerinden tahmin yürütmek yöntemin uygulanmasını kolaylaştırmaktadır (Demir, 2014).

Evet/hayır yöntemi uygulanış yönünden avantajlı olmasına rağmen kararlılık boyutu için aynı durum söz konusu değildir. Chinn ve Hertz (2002)'in klasik Angoff yöntemi ile evet/hayır yönteminin kararlılık durumları üzerine yaptıkları araştırmada, klasik Angoff yönteminin evet/hayır yöntemine göre daha kararlı olduğu saptanmıştır.

Sınır grup (borderline group) yöntemi. Sınır grup yönteminin temeli, testi alan bireylerden sınır gruptaki öğrencilerin belirlenmesi ve bu öğrencilerin test puanları yardımıyla kesme puanının oluşturulmasıdır (Kane, 1998). Bir başka öğrenci merkezli yöntem olan zıt gruplar yönteminde tüm öğrenci grubu yargıcılar tarafından “yeterli” ve “yetersiz” şeklinde iki gruba ayrılmaktadır. Ancak; bu işlem her zaman planlandığı kadar kolay olmamaktadır. Sınır grup yöntemi, bu ayrımı yapmak yerine yeterli ve yetersiz gruptaki öğrencilerin tüm grup içerisinde belirlenmesini, bu iki gruba dâhil olmayan öğrencilerin oluşturduğu grubun ise “sınır grup” olarak tanımlanmasını önermiştir. Sınır grup yönteminde tipik olarak uygulanan işlem, öğrencileri “yeterli”, “yetersiz” ve “sınır grup” olmak üzere 3 farklı gruba ayırmaktır (Cizek ve Bunch, 2007). Uygulanan test sonucu sınır gruptaki öğrencilerin test puanlarının medyanı veya ortalaması, testin kesme puanı olarak belirlenmektedir. Bir yargıcı tarafından sınır grup yöntemine göre kesme puanının elde edilmesi Tablo 3'te örneklendirilmiştir.

Tablo 3. Sınır grup yöntemi uygulama örneği.

Öğrenciler	Yeterli	Yetersiz	Test Puanları
Ö.1	✓		70
Ö.2		✓	35
Ö.3	✓		95
Ö.4			80
Ö.5	✓		85
Ö.6			60
Ö.7		✓	40
Ö.8			55
Ö.9		✓	25
Ö.10		✓	35
Ö.11	✓		75
Ö.12			70
Ö.13		✓	50
Ö.14			60
Ö.15		✓	45

Tablo 3 incelendiğinde, bir yargıcı 15 öğrencinin 4'ünü yeterli, 6'sını yetersiz kategorisinde sınıflandırmıştır. Bu gruptan “yeterli” veya “yetersiz” kategorisine girmeyen öğrenciler Ö.4, Ö.6, Ö.8, Ö.12 ve Ö.14'tür. Bu öğrenciler, sınır gruptaki öğrencileri oluşturmaktadır. Sınır gruptaki öğrencilerin test puanları sıralı bir dizi haline getirildiğine 55, 60, 60, 70 ve 80 dizilimi elde edilmektedir. Bu puanların medyan değeri 60'tır. 60 değeri bir yargıcı tarafından belirlenen testin kesme puanına karşılık gelmektedir. Testin kesme puanı ise tüm yargıcıların kesme puanlarının medyan değeridir.

Sınır grup yöntemi ile kesme puanı hesaplanırken sınır gruptaki öğrencilerin test puanlarının hem aritmetik ortalaması hem de medyan değeri kesme puanı olarak kullanılabilir. Aritmetik ortalama kullanıldığı durumlarda kesme puanı oldukça düşük veya oldukça yüksek verilerden etkilenebilir. Bu durum kesme puanının hatalı hesaplanmasına neden olmaktadır. Sınır gruptaki öğrenci sayısının az olması kesme puanındaki hata miktarını daha da arttırabilir. Medyan değerinin kullanılması durumunda ise medyanın uç değerlerden etkilenmemesi nedeniyle, hata miktarı azalmaktadır (Zieky vd., 2006).

Sınır grup yönteminde yargıcıların öğrenci grubunun bilgi, yetenek ve beceri düzeylerini iyi bir biçimde tanıması gereklidir (Taşdelen, 2009). Öğrenci merkezli yöntemlerin temelini bu varsayım oluşturmaktadır. Öğrenci grubunun özelliklerini yargıcıların iyi tanımaması durumunda “sınır grubu” oluşturan öğrenciler hatalı belirlenecek ve kesme puanı da hatalı hesaplanacaktır. Bu olumsuzluk, öğrenci grubunu iyi tanıyan öğretmenlerin yargıcı olarak seçilmesi ile giderilebilmektedir (Tülübaş, 2009).

Sınır grup yöntemi uygulanırken, dikkat edilmesi gereken konulardan biri de sınır gruptaki öğrencilerin test puanlarının dağılımıdır. Bu dağılım iki farklı şekilde olabilmektedir. Bu dağılımlardan birincisi verilerin kümelenmiş dağılımıdır. Kümelenmiş dağılım, yöntemin iyi çalıştığına işaret etmektedir. İkinci dağılım ise verilerin geniş bir dağılımıdır. Bu durum yöntemin iyi çalışmadığına işaret etmektedir. Geniş dağılımın gözlenme nedeni, yargıcıların belirlediği sınır gruptaki öğrencilerin “yeterli” veya “yetersiz” grupta yer almasından, yargıcıların sınır gruptaki öğrencileri belirlerken testin ölçtüğü kapsam dışında ölçütleri kullanmasından, yargıcıların değerlendirme esnasında belirledikleri kendilerine ait iç ölçütlerin yeterince tutarlı olmamasından kaynaklanabilmektedir (Zieky vd., 2006).

Standart belirleme yöntemlerinin karşılaştırıldığı araştırmalar. Standart belirleme yöntemleri ile ilgili alanyazın incelendiğinde, diğer yöntemlerle karşılaştırma amaçlı Angoff yönteminin daha fazla kullanıldığı, sınır grup yönteminin ise daha az kullanıldığı saptanmıştır. Bu araştırmalarda Angoff yöntemi, Nedelsky yöntemi (Cross Impara, Frary, ve Jaeger, 1984; Demir, 2014; Kara ve Kelecioğlu, 2015; Ömür ve Selvi, 2010; Taşdelen, 2009; Tanrıverdi, 2006), evet/hayır yöntemi (Çetin ve Gelbal, 2010; Gündeğer, 2012; Tanrıverdi, 2006), Jaeger yöntemi (Cross, Impara, Frary, ve Jaeger, 1984), Ebel yöntemi (Demir, 2014; Gündeğer, 2012; Ömür ve Selvi, 2010), işaretleme yöntemi (Çetin, 2011; Hsieh, 2013) ve sınır grup yöntemi (Shin, 2014; Tülübaş, 2009) ile karşılaştırılmıştır. Angoff yöntemi ile daha çok karşılaştırılan yöntemler ise Nedelsky ve evet/hayır yöntemleridir. Angoff ile evet/hayır yöntemlerinin karşılaştırıldığı çalışmalarda genellikle evet/hayır yöntemi ile elde edilen kesme puanının Angoff yöntemine göre elde edilen kesme puanından daha yüksek olduğu belirlenmiştir. Sınır

grup yöntemi ile Angoff yönteminden elde edilen kesme puanları karşılaştırıldığı araştırmalar incelendiğinde ise bir genellemeden bahsetmek zordur. Bazı araştırmalarda Angoff yönteminden, bazı araştırmalarda ise sınır grup yönteminden elde edilen kesme puanları diğerine göre daha yüksektir. Bu nedenle sınır grup yöntemi ile diğer yöntemlerin karşılaştırıldığı daha fazla sayıda araştırmaya ihtiyaç duyulmaktadır. Ayrıca ilgili alanyazın incelendiğinde; sınır grup, Angoff ve evet/hayır yöntemlerinin bir arada karşılaştırıldığı araştırmaların sınırlı olduğu belirlenmiştir. Bu bağlamda; araştırmanın konusunu Angoff, evet/hayır ve sınır grup yöntemleri ile elde edilen kesme puanlarının karşılaştırılması oluşturmaktadır.

Yöntem

Çalışma Grubu

Araştırmanın amacına uygun olarak 2 farklı gruptan veriler elde edilmiştir. İlk grup teste katılan öğrencilerdir. Öğrenci grubunu, 2017-2018 eğitim-öğretim yılında Uşak ilindeki merkeze bağlı 7 Anadolu lisesinde öğrenim gören 806 öğrenci oluşturmaktadır. Öğrenci grubu amaçsal örnekleme yöntemi ile gönüllülük esasına uygun olarak belirlenmiştir. Bu araştırmada sınır grup yönteminde yargıcıların öğrencilerle ilgili yeterli yetersiz kararı verebilmeleri için öğrencileri tanımaları gerekmektedir. Ayrıca 10. sınıf düzeyinde Türkçe ve Matematik testleri hazırlandığından 10. sınıfa devam eden öğrencilerin çalışma grubuna dahil edilmelidir. Bu nedenlerle yargıcı grubundan eğitim alan ve 10. sınıfa devam eden öğrencilerin çalışma grubuna dahil edilmesi amaçlanmıştır.

Araştırma kapsamında veri toplanan ikinci grup ise yargıcı grubudur. Yargıcı grubu, nihai teste katılan öğrenci grubunun matematik ve Türk dili ve edebiyatı dersine giren öğretmenlerden oluşmaktadır. Özellikle standart belirleme yöntemlerinden öğrenci merkezli yöntemlerde, araştırmaya katılan yargıcı grubunun öğrencilerin başarı düzeyini iyi bir şekilde tanınması gerekmektedir (Taşdelen, 2009; Tülübaş, 2009). Bu nedenle öğrencilerin dersine giren, 22 matematik ve 16 Türk dili ve edebiyatı öğretmeni yargıcı grubu olarak belirlenmiştir. Yargıcı grubundaki öğretmenler en az üç yıllık öğretmenlik deneyimine sahiptir.

Standart belirleme çalışmalarında görev alacak yargıcı sayısı için farklı görüşler bulunmaktadır. Kullanılan yönteme bağlı olarak yargıcı sayısı değişkenlik gösterse de bu konuda yapılan araştırmalarda genellikle yargıcı sayısının 10-15 arasında olması önerilmektedir (Hurtz ve Hertz, 1999; Taşdelen, 2009). Bazı çalışmalarda da yargıcı sayısı için ulaşılabilen ve çalışmaya gönüllü olarak katılan kişi sayısı ifadesi yer almaktadır (Tülübaş, 2009). Jeager (1989)'a göre ise yargıcı sayısı arttıkça araştırmanın standart hatası da azalmaktadır. Bu nedenle, araştırmaya gönüllü olarak katılan ve öğrenci grubunun derslerine giren matematik ve Türk dili edebiyatı öğretmenlerinin tamamı yargıcı olarak belirlenmiştir.

Veri toplama araçları ve verilerin toplanması. Araştırma kapsamında öğrencilerden veri toplamak için matematik ve Türkçe başarı testi ile yargıcılardan veri toplamak için yargıcı görüşleri belirleme formları oluşturulmuştur. Bu iki çalışma grubundan veri toplamak için oluşturulan veri toplama araçlarına ilişkin bilgiler ve bu araçlarla verilerin elde edilmesine yönelik süreçler iki ayrı başlıkta açıklanmıştır.

Matematik ve Türkçe başarı testleri. Öğrencilerin matematik ve Türkçe başarı düzeylerini belirlemek için test geliştirme çalışması yapılmıştır. Test geliştirme çalışmasına, hedeflerin ve konunun belirlenmesi ile başlanmıştır. Araştırmanın yapılacağı grubun 10. sınıf öğrencileri olması ve matematik, Türkçe dersleri için kesme puanlarına ulaşılacak istenilmesi nedeni ile 10. sınıf matematik ve Türk dili ve edebiyatı derslerinin öğretim programları incelenmiştir. Alan uzmanları ile yapılan değerlendirmeler sonucu matematik dersi için olasılık konusu, Türkçe dersi için ise isim ve sıfatlar konusunda test geliştirme çalışmasının yapılmasına karar verilmiştir. Bu kapsamda öğretim programlarının kazanımları, geliştirilmesi planlanan testin kazanımları olarak belirlenmiştir.

Kazanım belirleme işlemi tamamlandıktan sonra olasılık, isim ve sıfatlar konularına ilişkin belirtke tabloları hazırlanmıştır. Deneme testi için genelde nihai testte kullanılacak madde sayısının 3 katı kadar madde yazılması önerilmektedir. Ancak, testin uygulanacağı grubun özelliklerinin iyi tanınması halinde bu sayının azalabileceği Baykul (2015) tarafından belirtilmiştir. Testlerin uygulama süresinin bir ders saatini aşmaması da göz önünde bulundurularak deneme testlerinin toplam 40'ar maddeden oluşturulması planlanmıştır.

İlgili alan uzmanları ile matematik ve Türkçe deneme testleri için 40'ar madde yazılmıştır. Madde yazımı tamamlandıktan sonra kapsam geçerliği bakımından testlerin incelenmesi için kapsam geçerliği alan uzmanı görüş formu oluşturulmuştur. Bu formlar, alanlarında deneyimli 3 matematik ve 3 Türk dili ve edebiyatı öğretmeni tarafından incelenmiştir. Alan uzmanları tarafından yapılan incelemeler sonucunda deneme testlerinin kapsam bakımından uygun olduğu saptanmıştır. Ancak, bazı maddelerde düzeltmelerin yapılması alan uzmanları tarafından önerilmiştir. Alan uzmanlarının bu dönütleri sonrası deneme testlerinde gerekli düzeltmeler yapılmıştır.

Deneme testlerinin kapsam bakımından incelenmesi ve gerekli düzeltmelerin tamamlanması sonrasında, dil uygunluğu bakımından incelenmesi için araştırmacılar tarafından dil uzmanı görüş formları oluşturulmuştur. Bu formların incelenebilmesi için Türkçe alanında yüksek lisans mezunu olan iki dil uzmanına formlar elektronik ortamda gönderilmiştir. Dil uzmanlarından alınan görüşler doğrultusunda testlerde gerekli düzenlemeler yapılmıştır. Son olarak testlerin ölçme ve değerlendirme bakımından uygunluğunun incelenmesi için ölçme ve değerlendirme uzmanı görüş formları oluşturulmuştur. Bu formlar, ölçme ve değerlendirme alanında doktora öğrenimi görmekte olan bir uzmana elektronik ortamda gönderilmiştir. Ölçme ve değerlendirme uzmanının incelemesi tamamlandıktan sonra testlerde gerekli düzenlemeler yapılarak matematik ve Türkçe deneme testleri oluşturulmuştur.

40'ar maddeden oluşan deneme testlerinin uygulaması, çalışma grubunda bulunan okullardan rastgele belirlenen iki okulda, gönüllü 251 öğrenci ile gerçekleştirilmiştir. Her bir test için bir ders saati süre verilmiştir. Deneme testleri uygulanırken araştırmanın amacı ve önemi öğrencilere aktarılmış, sınavda kopya çekilmemesi için gerekli önlemler alınmıştır. Deneme uygulamasından elde edilen verilerin, test ve madde analizleri için Test Analysis Program (TAP) kullanılmıştır. Öğrencilere ait veriler TAP programına işlenirken öğrencilere sınav esnasında oturma düzenlerine göre verilen kodlar kullanılmıştır.

Ayırt edicilik indeksi madde kalitesi ile ilgili bilgi veren bir indekstir ve bu indeksin 0,40 ve üzerinde değerler alması maddenin ayırt edici bir madde olduğunun

göstergesidir (Crocker ve Algina, 1986). Bu nedenle TAP programında yapılan analizler sonucu ayırt edicilik indeksi 0,40 altında olan maddelerin nihai testten çıkarılması kararı alınmıştır. Bu kapsamda; matematik testinden 21, Türkçe testinden ise 25 madde testten çıkarılmıştır. Nihai testlerin bir ders saat içerisinde uygulanması ve kazanımlar için belirlenen madde sayıları da göz önünde bulundurularak nihai testin 30 maddelik tek bir formdan oluşturulması planlanmıştır. 15'i matematik ve 15'i Türkçe olmak üzere toplam 30 madde, madde güçlük indekslerine göre yeniden sıralanmış ve nihai test oluşturulmuştur.

Nihai testler, çalışma grubunda yer alan 5 okulda toplam 806 gönüllü öğrenciye uygulanmıştır. Uygulama öncesi okul idareleri ile görüşülerek araştırma hakkında bilgi verilmiştir. Okul idarelerinin uygun gördüğü zamanlarda testler genellikle ortak sınav şeklinde uygulanmıştır. Öğrencilere testleri cevaplayabilmeleri için bir ders saati süre verilmiştir. Testlerin uygulanması sürecinde öğrenciler araştırma hakkında bilgilendirilmiştir.

Matematik ve Türkçe dersleri için nihai testlerin madde ayırt edicilik indeksleri incelendiğinde, Türkçe testinin 10. maddesi hariç testte yer alan tüm maddelerin ayırt edicilik indekslerinin 0,40 değerinden büyük olduğu gözlemlenmiştir ($r_{jk} > 0,40$). Crocker ve Algina (1986)'ya göre bir madde için ayırt edicilik indeksinin 0,30-0,39 arasında olması durumunda, bu maddede küçük düzeltmeler yapılabilir veya hiçbir değişiklik yapılmadan da doğrudan teste alınabilir. Bu düşünceden hareketle Türkçe testinin 10. maddesi testten çıkarılmamıştır. Testlerin aritmetik ortalaması matematik testi için 66,7; Türkçe testi için ise 60,0 olarak hesaplanmıştır. Testlerin ortanca ve standart sapma değerleri sırasıyla matematik testi için 66,7 ve 4,15; Türkçe testi için ise 60,0 ve 3,88'dir. Her iki testin çarpıklık katsayıları incelendiğinde negatif değerler (Türkçe testi için -0,41, Matematik testi için -0,13) aldığı gözlemlenmektedir. Bu değerler testlerin her ikisinin de sola çarpık dağılım gösterdiğini işaret etmektedir. Testler için basıklık katsayıları negatif değerlere sahiptir (Türkçe testi için -1,11, Matematik testi için -1,16), testler için bu değerlerin negatif olmasının anlamı testlerin normalden daha basık bir dağılıma sahip olduğunun göstergesidir.

Sınıf içi başarı testlerinde, iç tutarlılık katsayısının en az 0,75 değerine sahip olması gerekmektedir (Murphy ve Davidshofer, 1991). KR-20 değerleri matematik için 0,86; Türkçe için 0,82 olarak hesaplanmıştır. Buna göre testler yeterli güvenilirlik düzeyine sahiptir. Ayrıca, testlerin ortalama madde ayırt edicilik indeksleri matematik için 0,64 ve Türkçe için ise 0,60 olarak hesaplanmıştır. Bu değerler testlerin iyi düzeyde ayrıncı olduğunu göstermektedir.

Yargıcı formları. Araştırma kapsamında matematik ve Türkçe grubu olmak üzere iki adet yargıcı grubundan veriler toplanmıştır. Bu nedenle her iki grup için ayrı form oluşturulmuştur. Formlar geliştirilirken nihai testlere ait maddeler ve bu testlere ait belirtke tabloları kullanılmıştır.

Yargıcılara karar verme süreci öncesinde araştırmanın amacı, araştırmada kullanılan yöntemler, performans düzeyleri, minimum yeterlik düzeyi, sınır grup, kesme puanı hakkında bilgilendirme toplantıları yapılmıştır. Bu toplantılar yargıcıların uygunluk durumuna göre bazen birebir bazen ise gruplar halinde gerçekleştirilmiştir. Yargıcı görüş formlarının yanıtlanmasına yönelik bilgiler bu toplantılarda yargıcılara aktarılmıştır. 3 farklı yöntemle göre karar alma süreçlerine yönelik bilgiler aşağıda açıklanmıştır.

Angoff yöntemine göre kesme puanı belirlenirken yargıcılardan öncelikle formda yer alan testteki 1. maddeyi incelemeleri istenilmiştir. Maddenin incelenmesinden sonra yargıcılara “geçme-kalma sınırında (minimum yeterlik düzeyinde) olan öğrencilerin yüzde kaç bu maddeyi doğru cevaplayabilir?” sorusu sorulmuştur. Yargıcılardan bu soruya yönelik bir tahminde bulunmaları ve bu tahminlerini formda yer alan ilgili kutucuğa yazmaları istenilmiştir. Geri kalan tüm maddeler için bu işlemin tekrarlanmasıyla birlikte Angoff yöntemi ile karar verme süreci tamamlanmıştır.

Evet/hayır yöntemine göre kesme puanı belirlenirken yargıcılardan öncelikle formda yer alan testteki 1. maddeyi tekrar incelemeleri istenilmiştir. Maddenin incelenmesinden sonra yargıcılara “geçme-kalma sınırında (minimum yeterlik düzeyinde) olan herhangi bir öğrenci bu maddeyi doğru cevaplayabilir mi?” sorusu sorulmuştur. Yargıcılardan bu soruya yönelik bir tahminde bulunmaları ve bu tahminlerini formda yer alan evet/hayır kutucuklarından birini işaretleyerek belirtmeleri istenilmiştir. Geri kalan tüm maddeler için bu işlemin tekrarlanmasıyla birlikte evet/hayır yöntemi karar verme süreci ve yargıcı görüşleri belirleme formunun yanıtlanması tamamlanmıştır.

Sınır grup yöntemi için yargıcıların belirli kriterlere göre grubu “yeterli”, “yetersiz” ve “sınır grup” olmak üzere 3 farklı gruba ayırması gerekmektedir. Bu ayırımın yapılabilmesi için kriter olarak öğretim programlarında yer alan kazanımlar, belirtke tablosu içinde yargıcılara verilmiştir. Belirtke tablosunda yer alan kazanımlara göre yargıcılardan yeterli olarak nitelendirdikleri öğrencileri sınıf listelerinde (+) işareti, yetersiz olarak nitelendirdikleri öğrencileri (-) işareti ile belirtmesi, bu iki grupta da yer almayan öğrenciler için ise herhangi bir işaretleme yapılmaması istenilmiştir. Bu işlemin tamamlanmasıyla birlikte sınır grup yöntemi ile karar verme süreci tamamlanmıştır.

Verilerin analizi. Araştırmanın birinci problemi olan “Angoff, evet/hayır ve sınır grup yöntemlerine göre kestirilen kesme puanları, bu kesme puanlarına göre belirlenen başarılı-başarısız öğrenci sayıları ve yüzdeleri nedir?” sorusunu yanıtlamak için 3 yönetime yönelik farklı kesme puanı hesaplama teknikleri kullanılmıştır.

Angoff yöntemi için minimum yeterlilik düzeyindeki öğrencilerin maddeye doğru yanıt verme olasılıklarının bir yargıcı için toplanması ve toplam madde sayısı olan 15’e bölünmesi sonucu o yargıcıya ait kesme puanı; tüm yargıcılar için belirlenen kesme puanlarının ortalamasının hesaplanması sonucu ise Angoff yöntemine göre testin kesme puanı elde edilmiştir.

Evet/hayır yöntemi için; bir yargıcının minimum yeterlilik düzeyindeki öğrencilerin bir maddeye doğru yanıt verme olasılıkları evet için “1”, hayır içinse “0” şeklinde kodlanmıştır. Bir yargıcıya ait tüm “1” değerlerin toplanması ve 15’e bölünmesi sonucu elde edilen değer 100 ile çarpılması ile o yargıcıya ait kesme puanı, tüm yargıcılar için belirlenen kesme puanlarının ortalamasının hesaplanması sonucu ise testin kesme puanı elde edilmiştir.

Sınır grup yönteminde ise yargıcıların yargıcı görüşleri belirleme formunda yer alan sınıf listelerinde herhangi bir işaretlemede bulunmayan öğrenciler, sınır grup öğrencilerini oluşturmuştur. Bir yargıcının belirlediği sınır gruptaki öğrencilerin test puanlarının ortancasının hesaplanması sonucu o yargıcıya ait kesme puanı, tüm

yargıcıların belirlediği sınır grup öğrencilerinin test puanlarının ortancasının hesaplanması sonucu ise testin kesme puanı elde edilmiştir.

Birinci problemin incelenmesine yönelik olarak 3 yöntem ile her bir ders için elde edilen kesme puanları ve test puanları Excel programına aktarılmıştır. Bu programda oluşturulan formüller yardımı ile başarılı-başarısız öğrenci sayıları ve yüzdeleri belirlenmiştir.

Araştırmanın ikinci problemi olan “Angoff, evet/hayır ve sınır grup yöntemlerine göre başarılı kabul edilen öğrenci oranları arasında anlamlı bir fark var mıdır?” sorusu için verilerin analizinde bağımlı iki oran veya yüzde arasındaki farkın belirlenmesinde z testi kullanılmıştır. Bu analiz, farklı yöntemlere göre başarılı kabul edilen ve edilmeyen kişilerin, toplam kişilere oranı ile hesaplanmaktadır (Korkmaz, 2015). Analiz sürecinde 2x2’lik tablolardan yararlanılmaktadır. Bu tablolarda yer alan değişkenlerin anlamı ve analiz yöntemi aşağıda açıklanmıştır.

Tablo 4. Matematik dersi Angoff, evet/hayır ve sınır grup yöntemleri için kesme puanları, başarılı-başarısız öğrenci sayıları ve yüzdeleri.

		<u>Yöntem 2</u>		
		Başarısız	Başarılı	
<u>Yöntem 1</u>	Başarılı	a	b	p ₁
	Başarısız	c	d	q ₁
		q ₂	p ₂	1.00

- a: Yöntem-1’den başarılı, Yöntem-2’den başarısız olan öğrenci oranı
b: Yöntem-1 ve Yöntem-2’den başarılı olan öğrenci oranı
c: Her iki yöntemden de başarısız olan öğrenci oranı
d: Yöntem-1’den başarısız, Yöntem-2’den başarılı olan öğrenci oranı
p₁: Yöntem-1’den başarılı olan öğrenci oranı
p₂: Yöntem-2’den başarılı olan öğrenci oranı
q₁: Yöntem-1’den başarısız olan öğrenci oranı
q₂: Yöntem-2’den başarısız olan öğrenci oranı

Bu değişkenler yardımı ile bağımlı iki oran arasındaki fark, z istatistiği ile hesaplanabilmektedir. Bu istatistik değeri; her iki yönteme göre başarılı olan öğrenci oranları arasındaki farkın, o oranlara ait standart sapmaya bölünmesi ile elde edilmektedir (Ferguson, 1959). z istatistiğine ait standart sapma ve z değeri aşağıdaki eşitlikler yardımı ile elde edilmektedir.

$$\sigma = \sqrt{\frac{a+d}{N}} \quad (\text{Standart Sapma})$$

$$z = \frac{p_1 - p_2}{s} \quad (\text{z değeri})$$

z istatistiğini elde edebilmek için öncelikle 3 standart belirleme yöntemi ile matematik ve Türkçe testlerinde başarılı ve başarısız öğrenci oranları Excel

programında hesaplanmıştır. Başarılı ve başarısız öğrenci oranları kullanılarak Excel programında oluşturulan formüller yardımı ile z değerleri hesaplanmıştır.

Bulgular

Araştırmanın birinci alt probleminin çözümüne yönelik öncelikli olarak 22 matematik ve 16 Türk dili ve edebiyatı dersi öğretmeni olan toplam 38 yargıcının verdiği kararlar doğrultusunda kesme puanları hesaplanmıştır. Tüm yargıcılardan elde edilen kesme puanları incelendiğinde; matematik dersi için en yüksek kesme puanının evet/hayır yöntemine göre 86,67; en düşük kesme puanının ise sınır grup yöntemine göre 16,67 olarak hesaplandığı gözlemlenmiştir. Türkçe dersi için ise; en yüksek kesme puanının Angoff yöntemine göre 93,67 olduğu saptanmıştır. Bu değeri, evet/hayır yöntemine göre elde edilen 93,33 değeri takip etmektedir. Türkçe dersi için Angoff ve evet/hayır yöntemlerine göre belirlenen en yüksek kesme puan değerleri birbirine oldukça yakındır. Türkçe dersi için en düşük kesme puanı ise sınır grup yöntemine göre 26,67 olarak hesaplanmıştır. Tüm yargıcılardan elde edilen kesme puanları kullanılarak Angoff, evet/hayır ve sınır grup yöntemlerine göre hesaplanan matematik ve Türkçe testlerine ait kesme puanları, başarılı-başarısız öğrenci sayı ve yüzdeleri Tablo 4 ve Tablo 5'te verilmiştir.

Tablo 5. Yöntemlere göre kesme puanları, başarılı-başarısız öğrenci sayıları ve yüzdeleri: Matematik dersi

Ders	Yöntem	Kesme Puanı	Başarılı Öğrenci		Başarısız Öğrenci	
			Sayı	Yüzde	Sayı	Yüzde
Matematik	Angoff	53,21	528	65,51	278	34,49
	Evet/hayır	54,82	481	59,68	325	40,32
	Sınır Grup	46,67	563	69,85	243	30,15

Tablo 4'teki matematik dersi için kesme puanları incelendiğinde; en yüksek kesme puanı evet/hayır yöntemine göre 54,82 olarak hesaplanmıştır. Bu değeri, Angoff yöntemine göre 53,21 değeri takip etmektedir. En düşük kesme puanının ise sınır grup yöntemine göre hesaplanan 46,67 olduğu gözlemlenmektedir. 806 öğrencinin test puanları ile yöntemler için elde edilen kesme puanları karşılaştırılması sonucu hesaplanan başarılı öğrenci sayıları ve yüzdeleri; Angoff yöntemi için 528 (%65,51), evet/hayır yöntemi için 481 (%59,68) ve sınır grup yöntemi için ise 563 (%69,85)'tür. Bu değerlere göre matematik testi için en yüksek başarı yüzdesi sınır grup yöntemi ile en düşük başarı yüzdesi ise evet/hayır yöntemi ile elde edilmektedir.

Tablo 6. Yöntemlere göre kesme puanları, başarılı-başarısız öğrenci sayıları ve yüzdeleri: Türkçe dersi

Ders	Yöntem	Kesme Puanı	Başarılı Öğrenci		Başarısız Öğrenci	
			Sayı	Yüzde	Sayı	Yüzde
Türkçe	Angoff	67,30	312	38,71	494	61,29
	Evet/hayır	63,33	376	46,65	430	53,35
	Sınır Grup	53,33	422	52,36	384	47,64

Tablo 7. Yöntemlere göre belirlenen kesme puanları, başarılı öğrenci sayıları, oranları ve z değerleri

Ders	Yöntemler	Kesme Puanı	Başarılı Öğrenci Sayısı	Başarılı Öğrenci Oranı	Z
Matematik	Angoff	53,21	528	0,66	
	Evet/hayır	54,82	481	0,60	6,85*
	Angoff	53,21	528	0,66	
	Sınır Grup	46,67	563	0,70	-5,91*
	Evet/hayır	54,82	481	0,60	
	Sınır Grup	46,67	563	0,70	-9,05*
Türkçe	Angoff	67,30	312	0,39	
	Evet/hayır	63,33	376	0,47	-8,00*
	Angoff	67,30	312	0,39	
	Sınır Grup	53,33	422	0,52	-10,48*
	Evet/hayır	63,33	376	0,47	
	Sınır Grup	53,33	422	0,52	-6,68*

Not. * $p < 0,01$ düzeyinde anlamlı

Tablo 5'teki Türkçe dersi için kesme puanları incelendiğinde; en yüksek kesme puanı Angoff yöntemine göre 67,30 olarak hesaplanmıştır. Bu değeri, evet/hayır yöntemine göre 63,33 değeri takip etmektedir. En düşük kesme puanının ise sınır grup yöntemine göre hesaplanan 53,33 olduğu gözlemlenmektedir. 806 öğrencinin test puanları ile yöntemler için elde edilen kesme puanları karşılaştırılması sonucu hesaplanan başarılı öğrenci sayıları ve yüzdeleri Angoff yöntemi için 312 (%38,71), evet/hayır yöntemi için 376 (%46,65) ve sınır grup yöntemi için ise 422 (%52,36)'dir. Bu değerlere göre Türkçe testi için en yüksek başarı yüzdesi sınır grup yöntemi ile en düşük başarı yüzdesi ise Angoff yöntemi ile elde edilmektedir.

İkinci araştırma problemine ilişkin olarak Angoff, evet/hayır ve sınır grup yöntemlerine göre başarılı kabul edilen öğrenci oranları arasında anlamlı bir farklılık

olup olmadığı belirlenmesi için z testi kullanılmıştır. Matematik ve Türkçe dersleri için elde edilen kesme puanları, başarılı öğrenci sayıları, oranları ve z değerleri Tablo 6'da verilmiştir.

Tablo 6 incelendiğinde; matematik ve Türkçe derslerinde Angoff, evet/hayır ve sınır grup yöntemlerine göre başarılı kabul edilen öğrenci oranlarının ikili karşılaştırılmalarının tamamı arasında 0,01 düzeyinde anlamlı fark bulunmaktadır.

Sonuç

Araştırma sonucu elde edilen bulgulara göre matematik ve Türkçe dersleri için Angoff, evet/hayır ve sınır grup yöntemlerine göre belirlenen kesme puanları birbirinden farklılaşmaktadır. Üç yöntem karşılaştırıldığında en yüksek kesme puanı matematik dersi için evet/hayır yöntemi ile; Türkçe dersi için ise Angoff yöntemi ile hesaplanmıştır. Bu bulgular matematik dersi için, Tanrıverdi (2006) ve Gündeğer (2012)'in Angoff ve evet/hayır yöntemlerini karşılaştırdıkları çalışmaların sonuçları ile uyum göstermekte, Çetin ve Gelbal tarafından yapılan araştırma sonuçlarıyla farklılaşmaktadır. Çetin ve Gelbal (2010) tarafından yapılan araştırmada Matematik dersi için evet/hayır yöntemine dayalı belirlenen kesme puanı Angoff yöntemine dayalı kesme puanından daha düşük bulunmuştur. Türkçe dersi için de kesme puanlarına ilişkin bulgular bu araştırmanın sonuçlarıyla farklılaşmıştır. Çetin ve Gelbal (2010) tarafından yapılan araştırmada Türkçe dersi için Angoff yöntemine dayalı belirlenen kesme puanı evet/hayır yöntemine dayalı kesme puanından daha düşük bulunmuştur. Araştırma sonucu her iki ders için en düşük kesme puanları sınır grup yöntemi ile hesaplanmıştır. Bu bulgu; Angoff ve sınır grup yöntemlerinin karşılaştırıldığı Tülübaş (2009) ve Shin (2014)'in araştırma sonuçları ile benzerlik göstermekte, Mills (1983)'in araştırma sonuçları ile örtüşmemektedir. Sınır grup yönteminden elde edilen kesme puanının diğer yöntemlere göre daha düşük olmasının birden çok nedeni olabilir. Bu nedenlerden birisi sınır gruptaki bireylerin özelliklerinin yeterince tanınmaması olabilir. Uygulamaya katılan okullardaki sınıf mevcutlarının kalabalık olması, öğrencilerin ölçülen özellik bakımından yeterince tanınmamasına neden olabilmektedir. Bir diğer neden yargıçların sınır gruptaki bireylerin bilgi ve becerilerini ölçüt olarak kullanmak yerine farklı ölçütleri işe koşarak kararlarını alması olabilir. Bir başka neden ise sınır gruptaki bireylerin, yargıçların beklediği performansa ulaşamamasından kaynaklanabilir.

Kesme puanlarına bağlı olarak, üç yönteme göre her iki ders için başarılı kabul edilen öğrenci oranları farklılaşmaktadır. Başarılı kabul edilen öğrenci oranları üç yöntem için karşılaştırıldığında en yüksek başarı oranları sınır grup yöntemine aittir. Bu bulgular, Angoff ve evet/hayır yöntemlerinin karşılaştırıldığı Tanrıverdi (2006) ve Gündeğer (2012)'in çalışmalarının sonuçları ile uyumludur. Ayrıca bu bulgular; Tülübaş (2009)'in Angoff ve sınır grup yöntemlerini karşılaştırdığı çalışmasıyla da örtüşmektedir. Yöntemlerin tamamı arasında başarılı öğrenci oranları arasındaki farklılık birçok nedenden kaynaklanabilir. Örneğin; yöntemlerin odaklandığı parametre bu nedenlerden birisidir. Angoff, evet/hayır yöntemlerinde odaklanılan parametre, testin içeriği ve test maddeleridir. Ancak, sınır grup yönteminde ise testi alan bireylerin bilgi ve becerileri hakkında yargıcı görüşleri temel parametre olarak kullanılmaktadır. Sınır grup ve diğer iki yöntem arasında farklılaşmanın kaynağı bu parametreler olabilir. Angoff ve evet/hayır yöntemleri arasında gözlemlenen farklılaşma ise iki farklı durumla

açıklanabilir. Bu durumlardan ilki yöntemlerin puanlama biçimidir. Angoff yönteminde yargıçlar test maddeleri hakkında karar verirken sürekli veri kullanmakta; evet/hayır yönteminde ise yargıçlar kategorik (0-1) verilerini kullanılmaktadır. Bu durumlardan ikincisi ise yargıçların Angoff yöntemine kıyasla evet/hayır yönteminde daha kolay karar vermeleridir (Impara ve Plake, 1997). Tüm bu belirtilen nedenlerden dolayı 3 yöntemle göre başarılı kabul edilen öğrenci oranları arasında anlamlı farklılaşma tespit edilmiş olabilir.

Öneriler

Bu araştırmanın sonuçlarından yararlanılarak standart belirleme uygulamalarına ve gelecekte yapılacak araştırmalara yönelik bir takım önerilerde bulunulabilir. Kesme puanının bir test için düşük olması istenildiği durumlarda öğrenci merkezli standart belirleme yöntemlerinin, yüksek olması istenildiği durumlarda ise test merkezli standart belirleme yöntemlerinin kullanılması uygulayıcılara önerilebilir.

Bu araştırmada, test merkezli standart belirleme yöntemlerinden Angoff ile evet/hayır yöntemleri; öğrenci merkezli standart belirleme yöntemlerinden ise sınır grup yöntemi kullanılmıştır. Benzer bir çalışma farklı yöntemlerle, örneğin sınır grup yöntemi yerine zıt gruplar yöntemi kullanılarak yapılabilir. Bu araştırmada kesme puanları belirlenirken matematik ve Türkçe testleri kullanılmıştır. Benzer bir çalışma farklı testler örneğin fen veya sosyal testleri kullanılarak da yapılabilir. Benzer bir çalışmada; standart belirleme süreçlerinde, test ve madde parametreleri karar verme süreci öncesi ve sonrası yargıçlara bildirilerek; yargıçların aldığı kararların karşılaştırılabilir.

Kaynakça

- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thomdike (Ed.), *Educational measurement* (2nd ed., pp. 508-600). Washington, DC: American Council on Education.
- Atılğan, H., Yurdakul, B. ve Öğretmen, T. (2012). Öğrenci başarısının belirlenmesinde bağıl ve mutlak değerlendirme üzerine bir araştırma. *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, 13(2), 79-98.
- Baykul, Y. (2015). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem Akademi.
- Berk, R. A. (1984). *A guide to criterion-referenced measurement: The state of the art: Baltimore, MD: Johns Hopkins University Press.*
- Brandon, P. R. (2004). Conclusions about frequently studied modified Angoff standard-setting topics. *Applied Measurement in Education*, 17(1), 59-88.
- Busch, J. C. ve Jaeger, R. M. (1990). Influence of type of judge, normative information, and discussion on standards recommended for the national teacher examinations. *Journal of Educational Measurement*, 27(2), 145-163.
- Chinn, R. N. ve Hertz, N. R. (2002). Alternative approaches to standard setting for licensing and certification examinations. *Applied Measurement in Education*, 15(1), 1-14.

- Cizek, G. J. (2001). Conjectures on the rise and fall of standard setting: An introduction to context and practice. Gregory J. Cizek (Ed.), *Setting performance standards: Theory and Application* (pp. 3-17). New Jersey: Lawrence Erlbaum Associates.
- Cizek, G. J. ve Bunch, M. (2007). *Standard setting, a guide to establishing and evaluating performance standards on tests*. Thousand Oaks, London, New Delhi: SAGE Publications.
- Clauser, B. E., Mee, J. ve Margolis, M. J. (2013). The effect of data format on integration of performance data into Angoff judgments. *International Journal of Testing*, 13 (1), 65-85.
- Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*: Newyork: Holt.Rinehart and Winston.
- Cross, L. H., Impara, J. C., Frary, R. B. ve Jaeger, R. M. (1984). A comparison of three methods for establishing minimum standards on the national teacher examinations. *Journal of Educational Measurement*, 21(2), 113-129.
- Çetin, S. ve Gelbal, S. (2010). Farklı standart belirleme yöntemlerinin geçme puanları üzerine etkisi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 43(1), 79-95.
- Çetin, S. (2011). *İşaretleme ve Angoff standart belirleme yöntemlerinin karşılaştırılması*. Yayımlanmamış yüksek lisans tezi, Hacettepe üniversitesi, Ankara.
- Demir, O. (2014). *Angoff, Nedelsky ve Ebel standart belirleme yöntemleri ile belirlenen kesme puanlarının karşılaştırılması*. Yayımlanmamış yüksek lisans tezi, Abant İzzet Baysal Üniversitesi, Bolu.
- Doğan, N. ve Gündeğer, C. (2014). Angoff, yes/no ve ebel standart belirleme yöntemlerinin karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 5(1),53-60.
- Ferguson, G. A. (1959). *Statistical analysis in psychology and education*. New York: McGraw-Hill.
- Gündeğer, C. (2012). *Angoff, Yes/No ve Ebel standart belirleme yöntemlerinin karşılaştırılması*. Yayımlanmamış yüksek lisans tezi, Hacettepe üniversitesi, Ankara.
- Hambleton, R.K. (2001). Setting performance standards on educational assessments and criteria for evaluating the process. Gregory J.Cizek (Ed.), *Setting performance standards: Theory and application* (pp. 89-116). New Jersey: Lawrence Erlbaum Associates.
- Hambleton, R. K. ve Pitoniak, M. J. (2006). Setting performance standards. *Educational Measurement*, 4(1), 433-470.
- Hsieh, M. (2013). Comparing yes/no Angoff and bookmark standard setting methods in the context of English assessment. *Language Assessment Quarterly*, 10(3), 331-350.
- Hurtz, G. M. ve Hertz, N. R. (1999). How many raters should be used for establishing cutoff scores with the Angoff method? A generalizability theory study. *Educational and Psychological Measurement*, 59(6), 885-897.
- Impara, J. C. ve Plake, B. S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34(4), 353-366.

- Jeager, R. (1989). Certification of student competence in educational measurement. R.L. In R. L. Linn (Ed.), *Educational measurement* (pp. 485-514). New York: Macmillan Publishing Co.
- Kane, M. (1998). Criterion bias in examine centered standard setting: some thought experiments. *Educational Measurement: Issues and Practice*, 17(1), 23-30.
- Kara, Y. ve Kelecioğlu, H. (2015). Puanlayıcı niteliklerinin kesme puanlarının belirlenmesine etkisinin genellenebilirlik kuramı'yla incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 58-71.
- Korkmaz, S. (2015). *Evet / Hayır, Ebel ve işaretleme standart belirleme yöntemlerinin karşılaştırılması*. Yayınlanmamış yüksek lisans tezi, Hacettepe üniversitesi, Ankara.
- Livingston, S. A. ve Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests*. New Jersey: Educational Testing Service.
- Murphy, K. R. ve Davidshofer, C. O. (1991). *Psychological testing: Principles and applications*. New Jersey: Prantice Hall.
- Norcini, J. J. (2003). Setting standards on educational tests. *Medical education*, 37(5), 464-469.
- Ömür, S., ve Selvi, H. (2010). Angoff, Ebel ve Nedelsky yöntemleriyle belirlenen kesme puanlarının sınıflama tutarlılıklarının karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 109-113.
- Plake, B. S. (1998). Setting performance standards for professional licensure and certification. *Applied Measurement in Education*, 11(1), 65-80.
- Raymond, M. R. ve Reid, J. B. (2001). Who made thee a judge? Selecting and training participants for standard setting. Gregory J.Cizek (Ed.), *Setting performance standards: Theory and application* (pp. 133-172). New Jersey: Lawrence Erlbaum Associates.
- Rotherham, A. J. (2006). Making the cut: How states set passing scores on standardized tests. *Education Sector*. 19 Aralık 2017 tarihinde <https://www.air.org/sites/default/files/publications/EXPCutScores.pdf> sitesinden alınmıştır.
- Shin, S.K. (2014). How good is good enough?: A comparison of three methods for establishing cut scores on placement tests. *English Teaching*, 69 (4), 53-75.
- Tanrıverdi, S. (2006). *Standart belirleme yöntemlerinin geçme puanları üzerine etkisi*. Yayınlanmamış yüksek lisans tezi, Hacettepe üniversitesi, Ankara.
- Taşdelen, G. (2009). *Nedelsky ve Angoff standart belirleme yöntemlerinin genellenebilirlik kuramı ile karşılaştırılmasına ilişkin bir araştırma*. Yayınlanmamış yüksek lisans tezi, Hacettepe Üniversitesi, Ankara.
- Taşdemir, F. (2013). *Angoff (1-0), Nedelsky ve sınır değerleri saptama yöntemleri ile bir testin sınıflama doğruluklarının incelenmesi*. Yayınlanmamış yüksek lisans tezi, Ankara Üniversitesi, Ankara.
- Tülübaş, G. (2009). *Psikolojik testlerde Angoff ve Sınır Grup Yöntemleri ile kesme puanlarının belirlenmesi*. Yayınlanmamış yüksek lisans tezi, Hacettepe üniversitesi, Ankara.
- Wyse, A. E. (2009). *A comprehensive item response theory framework for evaluating standard setting*. Yayınlanmamış doktora tezi. Retrieved from ProQuest Dissertations and Theses database.(Accession No. 3395454).

Zieky, M., Perie, M. ve Livingston, S. (2006). *A primer on setting cut scores on tests of educational achievement*. Princeton, NJ: Educational Testing Service.

Comparison of Cut Points According to Angoff, Yes/No and Borderline Group Methods

Abstract

This study aims to compare the cut-off points obtained by two test-centered standard setting methods, namely the Angoff and the yes/no methods with those obtained by a student-centered standard setting method, namely the borderline group method. The data were collected from two different study groups: students and judges. The former consisted of a purposive sample of 1.057 students studying at 10th grade in Anatolian High Schools in Uşak, whereas the latter consisted of a purposive sample of 22 mathematics and 16 Turkish language and literature teachers who taught the students that participated in the study. The findings point to significant differences among the methods in terms of the percentages of the students considered successful.

Keywords: Angoff, Yes/No, borderline group, cut-off score designs