

Nedelsky ve Angoff Standart Belirleme Yöntemleri ile Elde Edilen Kesme Puanlarının Genellenabilirlik Kuramı ile Karşılaştırılması

Gülşen TAŞDELEN*

Hülya KELECİOĞLU**

Neşe GÜLER***

Bursa İnegöl Altıeylül İ.Ö.O.

Hacettepe Üniversitesi

Sakarya Üniversitesi

Özet

Bu araştırmada Nedelsky ve Angoff standart belirleme yöntemleri ile elde edilen kesme puanlarının Genellenebilirlik (G) Kuramı'na göre hesaplanan varyans bileşenleri karşılaştırılmıştır. Araştırmanın verileri 2008 yılı 6. sınıf Seviye Belirleme Sınavı'nda yer alan fen ve teknoloji sorularının Ankara ilinde görev yapan 40 fen ve teknoloji öğretmeni tarafından standart belirleme yöntemlerine uygun şekilde puanlanması ile elde edilmiştir. Kesme puanlarının G kuramı uygulamasıyla varyans bileşenleri ile G ve Φ katsayıları elde edilmiştir. Araştırmanın sonunda, her iki yönteme göre ayrı ayrı yapılan puanlamanın maddeler arası farklılığı ortaya çıkarabildiği belirlenmiştir. Angoff yöntemi ile belirlenen kesme puanları için puanlayıcılar arası tutarlılık görülmüş; Nedelsky yönteminde ise puanlayıcılar arasındaki uyumun düşük olduğu ortaya çıkmıştır. Nedelsky yöntemi ile elde edilen G ve Φ katsayıları daha yüksektir. Her iki standart belirleme yönteminin birlikte ele alındığı G çalışmasında ise maddeler arası farklılıkların ortaya çıkarılabildiği ve puanlayıcılar arasında bir tutarlılığın olduğu görülmüştür.

Anahtar Sözcükler: standart belirleme, Nedelsky yöntemi, Angoff yöntemi, genellenebilirlik kuramı

Abstract

In this study, according to generalizability theory (G theory) the variance components of the cutting scores obtained from Nedelsky and Angoff standards setting methods were compared. The data of this research were collected by using the 16 science and technology items which were asked on elementary 6th grade "Achievement Determination Exam (Seviye Belirleme Sınavı) (SBS)". These items were graded by 40 science and technology experts (raters) working in Ankara according to these two standards setting methods. The variance components of the cutting scores were obtained by using G theory. According to the results of the study, it is determined that the differences of the items can be revealed by using both Angoff and Nedelsky standards setting methods. While there is a rater consistency for the cutting scores obtained by Angoff method, rater consistency was not observed for the cutting scores obtained by Nedelsky method. In G study which both methods were taken in hand together, it was observed that the differences of the items could be revealed and there is a rater consistency for the cutting scores.

Keywords: standards setting, Nedelsky method, Angoff method, generalizability theory

Standart belirleme, bireyler hakkında karar vermek veya sınıflandırma yapmak amacıyla performans düzeylerini kesme puanları ile belirleme olarak tanımlanmaktadır (Irwin, 2007; Crocker ve Algina, 1986; Cizek, 2001). Jaeger (1989), kesme puanı belirleme yöntemlerini test merkezli ve öğrenci merkezli yöntemler olarak ikiye ayırmıştır.

* Fen ve Teknoloji Öğretmeni, Bursa İnegöl Altıeylül İlköğretim Okulu, gulsentasdelen@yahoo.com

** Doç. Dr., Hacettepe Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, hulyaebb@hacettepe.edu.tr

*** Yrd. Doç. Dr., Sakarya Üniversitesi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, gnguler@gmail.com

Test merkezli yöntemler sınav soruları ile ilgili uzmanların ya da öğretmenlerin yargılarına dayanır. Bu yöntemde puanlayıcılar, testteki maddeleri gözden geçirerek belirli bir performans düzeyi için kesme puanı belirlemeye çalışırlar. Test merkezli yöntemler, Angoff, Değiştirilmiş Angoff, Ortalama Kestirme Yöntemi, Nedelsky, Ebel, Bookmark ve Mapmark yöntemleridir.

Öğrenci merkezli yöntemler testi alan bireyler ve bu bireylerin yetenekleri göz önünde bulundurularak uygulanır (Irwin, 2007: 12). Sıklıkla kullanılan iki öğrenci merkezli yöntem Sınır Grup Yöntemi ve Karşıt Gruplar Yöntemi'dir.

Bu araştırmada test merkezli yöntemlerden Angoff ve Nedelsky kullanılmıştır.

Angoff Yöntemi

Bu yöntemde alan uzmanlarından oluşan puanlayıcılardan minimum yeterli düzeyindeki öğrencilerin testte bulunan her maddeyi cevaplayabilme durumu hakkında yargıya varmaları istenir (Irwin, 2007; 14). Her puanlayıcının testte bulunan bütün maddelere ilişkin sonuçları toplanarak o puanlayıcının Minimum Geçme Puanı (MGP) hesaplanır. Puanlayıcıların her birine ait bireysel MGP'lerin ortalaması alınarak testin kesme puanı elde edilir.

Nedelsky Yöntemi

Nedelsky yöntemi, sadece çoktan seçmeli maddelerde uygulanabilir. Bu yöntemde puanlayıcılar, minimum yeterli düzeyindeki öğrencilerin test maddesindeki çeldiricilerden kaç tanesini eleyebileceklerine karar verirler (Akt. Hauger; Nedelsky, 1954). Her madde için minimum yeterli düzeyindeki öğrencinin eleyemeyeceği seçenek sayısının toplamı o maddenin puanı olarak alınır. MGP ve kesme puanı, Angoff yönteminde olduğu gibi belirlenir.

Angoff yönteminde uzmanlar sadece madde kökü ve doğru cevaba odaklanırken, Nedelsky yönteminde madde kökü ile birlikte tüm seçeneklere de odaklanmaktadır (Cizek ve Bunch, 2006).

Standart belirleme yöntemleri ile elde edilen kesme puanlarına pek çok kaynaktan hata karışabilmektedir. Kesme puanını belirleyen uzmanlar ya da örnekleme alınan maddeler birer hata kaynağı olabilmektedir (Lee ve Lewis, 2008). Kesme puanlarının standart hatasını, birden fazla hata varyansının aynı anda kestirilebildiği Genellenebilirlik kuramı (G) ile belirlemek mümkündür (Brennan ve Lockwood, 1980; Lee ve Lewis, 2008).

Genellenebilirlik Kuramı

Genellenebilirlik kuramı, ölçme sonuçlarının güvenilirlikleri ile ilgili bir kuramdır (Shavelson ve Webb, 1991). Bu kuram, araştırmacının bir ölçme sırasında ortaya çıkabilecek pek çok hata kaynağını aynı anda ele alabilmesini sağlar. Genellenebilirliğin kökleri klasik test kuramı'na (KTK) ve varyans analizine dayanır. Ancak genellenebilirlikteki kavramsal çerçeve daha geniştir (Brennan, 2002).

KTK'nda ölçme sonuçlarına karışan hatalar, sadece bir değişkenlik kaynağından gelen hatalar olarak ele alınır ve bu durum KTK'nın bir sınırlılığıdır. Ancak G kuramı ile birden fazla değişkenlik kaynağı aynı anda kestirilebilir.

Örneğin, test-tekrar test yöntemi ile kestirilen güvenilirlikte hata kaynağı *zamanken*; paralel formlar yöntemi ile kestirilen güvenilirlikteki hata kaynağı kullanılan formlar arasındaki *tutarlılıktır*. KTK'nın farklı hata kaynaklarını dikkate alan farklı yöntemlerle hesaplanan güvenilirlik katsayıları birbirinden farklı olabilirken; G kuramı tüm değişkenlik kaynaklarından gelebilecek hataların hepsini aynı anda ele alarak bir tek güvenilirlik katsayısı hesaplanmasını sağlamaktadır. G Kuramı ile hata kaynaklarının ayrı ayrı ele alınmasının yanı sıra, bunların etkileşimi de belirlenebilmektedir.

Genellenebilirlik kuramında güvenilirlik kavramı, araştırmacının genellemek istediği basit veya karmaşık her evrene uygulanabilir. En basit evren tek değişkenlik kaynaklı evren iken değişkenlik kaynağı sayısının artması evreni de karmaşık hale getirir (Shavelson ve Webb, 1991). G kuramında; çalışmanın amacına, verilerin toplanmasında uygulanan desene, değişkenlik kaynağının türüne, sonuçların nasıl yorumlanacağına ve elde edilecek güvenilirlik katsayısına bağlı olarak farklı desenler bulunmaktadır.

Bir faktörün bütün koşulları diğer değişkenlik kaynağının bütün koşulları ile gözlemlenirse bu tip ölçmeler çaprazlanmış (crossed) desen; bir değişkenlik kaynağının sadece bazı koşulları diğer değişkenlik kaynağının bazı koşullarınca gözlemlenirse yuvalanmış (nested) desen olarak adlandırılır (Shavelson ve Webb, 1991). Ayrıca bazı değişkenlik kaynaklarının çaprazlanmış bazılarının da yuvalanmış olduğu karışık (mixed) desenli çalışmalar da söz konusudur (Mushquash ve O'Connor, 2006).

G Kuramında değişkenlik kaynakları tesadüfi ya da sabit olarak ele alınır. Araştırmacı "Koşulları örneklemin ötesinde evrene genelleme yapmak istiyor muyum?" sorusuna *evet* cevabını veriyorsa değişkenlik kaynağı tesadüfi; *hayır* cevabını veriyorsa sabittir. Değişkenlik kaynağının tesadüfi veya sabit olması ölçmenin genellenebilirliğini etkilemektedir (Shavelson ve Webb, 1991; 12).

G kuramı ölçmelerde hataya sebep olan varyans kaynaklarına odaklanmasına rağmen, genellenebilirlik katsayısı adında bir güvenilirlik katsayısı da sağlamaktadır. G katsayısı, bir bireyin gözlenen puanından, o bireyin evren puanının ne derece doğru genellendiğinin bir göstergesidir. KTK'ndaki güvenilirlik katsayısı gibi, genellenebilirlik katsayısı da bireylerin puanlarındaki çeşitliliğin oranını yansıtır (Güler, 2008).

Bu araştırmada, 2008 yılında uygulanan Seviye Belirleme Sınavı'nın (SBS) Fen ve Teknoloji alt testinin Angoff ve Nedelsky yöntemleriyle belirlenen kesme puanlarının varyans bileşenleri G Kuramı ile incelenmiştir. Araştırmada, tek değişken kaynaklı çaprazlanmış model kullanılmıştır (*bxm*).

Yöntem

Bu araştırma, Angoff ve Nedelsky yöntemleri ile belirlenen kesme puanlarının varyans bileşenleri ile G ve Φ katsayıları incelediğinden temel araştırmadır.

Çalışma Grubu

Araştırmanın çalışma grubunu, 2008-2009 öğretim yılında Ankara ilinde 13 ilköğretim okulunda görev yapan 40 Fen ve Teknoloji öğretmeni ve 415 ilköğretim 6. sınıf öğrencisi oluşturmaktadır. Öğretmenler, 2008 SBS Fen ve Teknoloji sorularını Angoff ve Nedelsky standart belirleme yöntemlerine göre değerlendirmişlerdir. Bu okullardaki öğrencilerin SBS Fen ve Teknoloji alt testindeki doğru cevap sayıları veri olarak kullanılmıştır.

Verilerin Toplanması

Öğretmenlere 2008 yılı SBS'de yer alan fen ve teknoloji sorularının bulunduğu bir form verilmiştir. Öğretmenlerden, Angoff yöntemi için minimum yeterli düzeyindeki öğrencinin soruyu cevaplayıp cevaplayamama durumunu; Nedelsky yöntemi için ise minimum yeterli düzeyindeki öğrencinin her bir maddedeki eleyebileceği seçenek sayısını belirtmeleri istenmiştir.

Verilerin Analizi

Verilerin analizi iki aşamada gerçekleştirilmiştir. Birinci aşamada 40 puanlayıcının her biri için standart belirleme yöntemlerine göre iki kesme puanı hesaplanmıştır. Daha sonra bu puanların ortalamaları alınarak her bir yöntem için kesme puanı elde edilmiştir. İkinci aşamada da genellenebilirlik kuramına göre G çalışması yürütülerek ana ve ortak etkiler için varyans değerleri kestirilmiştir. Kesme puanları Microsoft Office Excel programında hesaplanmıştır. Varyans bileşenleri ile G ve Φ katsayıları SPSS için yazılmış olan bir syntax ile elde edilmiştir (Mushquash ve O'Connor 2006; Pallant, 2007).

Bulgular

Angoff yöntemi ile elde edilen kesme puanının varyans bileşenlerinin belirlenmesi: Angoff yöntemi ile belirlenen kesme puanının G çalışması ile elde edilen varyanslarını ve varyans yüzdelerini hesaplamak amacıyla tümüyle çaprazlanmış *m_{xp}* modeli uygulanmıştır. 16 madde (*m*) ve 40 puanlayıcı (*p*) ile tek değişkenli modelle yapılan G çalışması için *m* ve *p* ana etkileri ile *mp* ortak

etkisinin kestirilen varyans bileşenleri, toplam varyansı açıklama yüzdeleri ile G ve Φ katsayıları Tablo 1’de verilmiştir.

Tablo 1’de Angoff standart belirleme yöntemi ile belirlenen kesme puanlarında madde ana etkisinin toplam varyansın % 57’sini; puanlayıcı ana etkisinin %0’ını açıkladığı görülmektedir. Hesaplama sonucu negatif çıkan puanlayıcı ana etkisi 0.00 olarak verilmiştir. Brennan (2001), G kuramında kestirilen varyans bileşeninin negatif değerli olması durumunda, Cronbach’ın diğer varyans bileşenlerini hesaplamak için negatif olan varyans bileşeninin sıfır alınmasını önerdiğini belirtmektedir. Hem Brennan hem de Cronbach genellenebilirlik katsayısının hesaplanmasında kestirilen negatif varyans değeri yerine sıfır kullanılmasının uygun olduğunu vurgulamaktadırlar (Shavelson ve Webb, 1991; Brennan, 2001; Atılgan, 2004).

Tablo 1. Angoff Yöntemi için G çalışması ile ölçmenin kestirilen varyansları ve toplam varyansı açıklama oranları

Varyans Kaynağı	Sd	Toplam Kareler	Kareler Ortalaması	Varyans	%	G	Φ
Madde (m)	15	90.51	6.03	0.15	0.57		
Puanlayıcı (p)	39	3.05	0.79	0.00	0.00	0.981	0.981
mp	585	66.43	0.11	0.11	0.43		
Toplam					100		

Madde x puanlayıcı (mp) ortak etkisi ise toplam varyansın % 43’ünü açıklamaktadır. Angoff yöntemi ile yapılan standart belirleme çalışmasında maddeler için kestirilen varyans bileşeninin toplam varyans içinde en yüksek paya sahip olduğu, bunu *madde x puanlayıcı (mp)* ortak etkisinin izlediği görülmektedir. Puanlayıcılar arası varyansın toplam varyans içinde bir payı yoktur. Angoff standart belirleme yöntemine göre elde edilen G ve Φ katsayısı da 0.981 olarak kestirilmiştir. Elde edilen puanların G ve Φ katsayıları birbirine eşittir. G katsayısı, Cronbach alfa gibi yorumlanır (Sudweeks, Reeve ve Bradshaw, 2005; Güler, 2008). Bu durumda Angoff standart belirleme yöntemine göre elde edilen kesme puanlarının güvenilirliğinin oldukça yüksek olduğu söylenebilir.

Nedelsky yöntemi ile elde edilen kesme puanının varyans bileşenlerinin belirlenmesi: Nedelsky yöntemi ile belirlenen kesme puanının G çalışması ile varyanslarını ve varyans yüzdelerini elde etmek amacıyla tümüyle çaprazlanmış mxp modeli uygulanmıştır. Ölçmenin yapıldığı 16 madde (m) ve 40 puanlayıcı (p) için tek değişkenli modelle yapılan G çalışması için m ve p ana etkileri ile mp ortak etkilerinin kestirilen varyans bileşenleri, toplam varyansı açıklama yüzdeleri ile G ve Φ katsayıları Tablo 2’de verilmiştir.

Tablo 2. Nedelsky Yöntemi için G çalışması ile ölçmenin kestirilen varyansları ve toplam varyansı açıklama oranları

Varyans Kaynağı	Sd	Toplam Kareler	Kareler Ortalaması	Varyans	%	G	Φ
Madde (m)	15	430.24	28.68	0.76	0.79		
Puanlayıcı (p)	39	60.87	1.56	0.20	0.21	1.00	0.993
mp	585	-993.30	-1.70	0.00	0.00		
Toplam					100		

Tablo 2’de Nedelsky standart belirleme yöntemi ile belirlenen kesme puanlarında madde ana etkisinin toplam varyansın % 79’unu; puanlayıcı ana etkisinin ise % 21’ini açıkladığı görülmektedir. *Madde x puanlayıcı (mp)* ortak etkisi ise % 0’lık değeri ile toplam varyansın açıklanmasında rol oynamamaktadır. Burada da negatif çıkan varyans değeri sıfır olarak verilmiştir. Nedelsky yöntemi ile yapılan standart belirleme çalışmasında maddeler için kestirilen varyans bileşeninin toplam varyans

içinde en yüksek paya sahip olduğu, bunu puanlayıcılar arası varyansın izlediği görülmektedir. *Madde x puanlayıcı (mp)* ortak etkisinin toplam varyans içinde bir payı yoktur. Nedelsky standart belirleme yöntemine göre kesme puanının belirlenmesinin tek boyutu altında yer alan uygulamadaki 16 madde sayısına ve bu maddeleri puanlayan 40 puanlayıcıya göre elde edilen G katsayısı 1.00 ve Φ katsayısı da 0.993 olarak kestirilmiştir. Bu durumda Nedelsky standart belirleme yöntemine göre elde edilen kesme puanlarının güvenilirliğinin oldukça yüksek olduğu söylenebilir.

Angoff ve Nedelsky Yöntemleri ile Elde Edilen Kesme Puanlarının Varyans Bileşenlerinin Belirlenmesi. İki yöntemle elde edilen kesme puanlarının G çalışması ile elde edilen varyanslarını ve varyans yüzdelerini hesaplamak amacıyla, tümüyle çaprazlanmış *mpxy* modeli uygulanmıştır. Ölçmenin yapıldığı 16 madde, 40 puanlayıcı ve 2 standart belirleme yöntemi (*y*) kullanılarak tek değişkenli modellerle yapılan G çalışması için *m*, *p* ve *y* ana etkileri ile *mp*, *my*, *yp* ve *mpy* ortak etkilerinin kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri Tablo 3’de verilmiştir.

Tablo 3’e göre Angoff ve Nedelsky standart belirleme yöntemleri ile belirlenen kesme puanları için kestirilen toplam varyansı açıklama oranları incelendiğinde, madde ana etkisi toplam varyansın % 22’sini; yöntem ana etkisi % 35’ini açıklamaktadır. Puanlayıcı ana etkisinin ise % 0’lık değeri ile toplam varyansın açıklanmasında rol oynamadığı görülmektedir. *Madde x puanlayıcı* ortak etkisi toplam varyansın % 16’sını *madde x yöntem (my)* ortak etkisi % 7’sini; *yöntem x puanlayıcı (yp)* ortak etkisi de % 2’sini açıklamaktadır. *Madde x puanlayıcı x yöntem* (artık) ortak etkisinin de toplam varyansı açıklama oranı % 18 olarak görülmektedir.

Tablo 3. Angoff ve Nedelsky yöntemleri için G çalışması ile ölçmenin kestirilen varyansları ve toplam varyansı açıklama oranları

Varyans Kaynağı	Sd	Toplam Kareler	Kareler Ortalaması	Varyans	%
Madde (<i>m</i>)	15	455.97	30.40	0.32	0.22
Puanlayıcı (<i>p</i>)	39	34.03	0.87	0.00	0.00
Yöntem (<i>y</i>)	1	328.05	328.05	0.51	0.35
<i>mp</i>	585	418.84	0.72	0.23	0.16
<i>my</i>	15	64.78	4.32	0.10	0.07
<i>yp</i>	39	29.89	0.77	0.03	0.02
<i>mpy</i>	585	150.29	4.32	0.26	0.18
Toplam					100

Tartışma ve Yorum

Bu araştırmada, Angoff ve Nedelsky standart belirleme yöntemleri ile elde edilen kesme puanlarının varyans bileşenleri G kuramı ile incelenmiştir.

Literatüre bakıldığında Brennan ve Lockwood (1980) yaptıkları çalışmada Angoff ve Nedelsky standart belirleme yöntemlerini genellenebilirlik kuramını kullanarak karşılaştırmışlar ve bu iki yöntem ile elde edilen kesme puanlarının oldukça farklı olduğu sonucuna ulaşmışlardır. Bu farklılığın sebepleri olarak da hem bu iki yöntemin uygulanışındaki farklılıkları hem de puanlayıcıların minimum yeterlik düzeyini algılama farklılıklarını göstermişlerdir.

Chang (1996), Nedelsky ve Angoff standart belirleme yöntemlerini karşılaştırmış; Nedelsky Yöntemi ile Angoff Yöntemine göre daha düşük kesme puanları belirlendiğini ve uzmanlar arası uyuşmazlığın daha düşük düzeyde olduğunu belirlemiştir. Araştırmada yirmi iki mezun öğrenci uzman olarak görev almış ve daha önce almış oldukları araştırma teknikleri dersi ile ilgili dokuz maddeyi her iki yöntemle göre puanlamaları istenmiştir. Araştırma sonucunda uzmanlar arası uyuşmazlık Nedelsky Yönteminde daha düşük bulunmuş; ayrıca, Nedelsky Yöntemi ile belirlenen kesme puanları öğrencilerin gerçek puanlarına oldukça yakın iken Angoff Yöntemi ile belirlenen kesme puanının öğrencilerin gerçek performanslarından büyük oranda saptığı gözlenmiştir.

Bu çalışmada ise Angoff ve Nedelsky yöntemleri ile elde edilen kesme puanlarındaki *madde* varyans bileşeni *puanlayıcı* ve *madde x puanlayıcı* etkileşimine ait varyans bileşenlerinden daha yüksektir. Maddelerin toplam varyansı açıklama yüzdesi Angoff ile 0,57; Nedelsky ile 0,79'dur. Maddeler için kestirilen varyansın toplam varyans içindeki oranının yüksek olması ölçme sürecinde istenen bir durumdur. Bu durum, yöntemlerle elde edilen kesme puanlarının maddeler arası farklılığı ortaya çıkarabildiğini göstermektedir. Bu sonuç aynı zamanda puanlayıcıların, maddelerle ölçülen özelliklerin farklılığını puanlamada dikkate aldığına da bir göstergesi olarak yorumlanabilir. Nedelsky yöntemi ile elde edilen kesme puanları, maddelere arası farklılığı daha iyi yansıtabilmiştir.

Puanlayıcı etkisi için elde edilen varyans bileşenleri yöntemlere göre farklılık göstermiştir. Angoff yöntemi ile puanlayıcı ana etkisi için G çalışmasından kestirilen varyans bileşeninin değeri 0.00 ile % 0'dır. Angoff yönteminde puanlayıcıların tüm maddeler için yaptıkları puanlamalar arasında bir farklılık bulunmadığı, puanlamalar arasında bir tutarlılığın olduğu söylenebilir. Nedelsky yönteminde ise puanlayıcılar arası varyans, ikinci sıradaki varyans bileşenidir ve toplam değişkenliğin %21'ini açıklamaktadır. Bu yöntemle elde edilen puanlar arasında fark olduğu, puanlayıcılar arası tutarlılığın düşük olduğu söylenebilir.

Madde ve puanlayıcı etkileşiminin etkisi Angoff yönteminde toplam değişkenliğin %43'ünü açıklarken, Nedelsky yönteminde bu etkinin toplam varyans üzerinde etkisinin olmadığı görülmüştür (%0). Angoff yönteminde puanlayıcılar arasında varyans olmamasına rağmen *madde x puanlayıcı* etkisinin görülmesi, tüm puanlayıcıların bazı maddelerde katı bazı maddelerde de cömert puanlama yapmış olabileceğini göstermektedir. Nedelsky yönteminde ise, puanlayıcıların maddelerle etkileşimi söz konusu olmamıştır. Kesme puanlarının G ve Φ katsayılarına bakıldığında ise her iki katsayının Nedelsky yönteminde Angoff yönteminden daha yüksek olduğu görülmüştür.

Her iki yöntem birlikte ele alınarak yapılan G çalışmasında ise, toplam değişkenlik üzerindeki en fazla etkinin yöntemlerden kaynaklandığı görülmüştür (%35). Bu durum, yöntemlerle elde edilen kesme puanlarının farklı olduğunu göstermektedir. Puanların ranjındaki farklılıktan dolayı kesme puanlarında da farklılık görülmesi beklenen bir durumdur. Toplam değişkenlik üzerinde ikinci sıradaki etki, madde etkisidir (%22). Bunu *madde x puanlayıcı x yöntem* (%18), *madde x puanlayıcı* (%16), *madde x yöntem* (%07) ve *yöntem x puan* (%02) etkileşimi izlemektedir. İki yöntem birlikte ele alındığında puanlayıcı etkisinin toplam varyans üzerinde bir etkisi yoktur (%0). Yöntem etkisini de ele alan G çalışması sonucunda yöntemler arası farklılık ortaya çıkmıştır. Yapılan çalışmada madde etkisinin kesme puanlarındaki değişkenliğin bir kısmını açıkladığı, yani iki yöntemin de madde etkisine duyarlı olduğu görülmektedir. Kesme puanlarında *madde x puanlayıcı x yöntem* ve *madde x puanlayıcı* etkileşiminin etkisi görülmekle birlikte; *madde x yöntem* ve *yöntem x puan* etkileşiminin etkisi oldukça düşüktür.

Angoff ve Nedelsky yöntemleri ile elde edilen kesme puanları maddelerdeki değişkenlikten etkilenmekte, yani maddelerdeki değişimi kesme puanlarına yansıtabilmektedir. Bu oran, Nedelsky yönteminde daha fazladır. Ancak, puanlayıcılar arasındaki varyansın Nedelsky yönteminde yer alması bu yöntemin zayıf bir yönü olarak değerlendirilebilir. Güvenirliği göreceli olarak Nedelsky yönteminden düşük olmakla birlikte, kullanım kolaylığı da düşünüldüğünde kesme puanlarının Angoff yöntemi ile belirlenmesi daha uygun görülmektedir. Ayrıca farklı standart belirleme yöntemlerinin karşılaştırıldığı bir araştırmada uzman sayısının 10 olarak en baştan ele alınması araştırmacılara uzmanlara ulaşabilme açısından kolaylık sağlayabilir.

Kaynaklar

- Atılğan, H. (2004). *Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılmasına İlişkin Bir Araştırma*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi: Ankara.
- Brennan, R. L. (2001). *Generalizability Theory*. Springer-Verlag New York, Inc.
- Brennan, R. L. (2002). *An NCNE Instructional Module on Generalizability Theory*. American College Testing.
- Brennan, R. L., & Lockwood, R. E. (1980). A comparison of two cutting score procedures. *Applied*

Psychological Measurement, 4, 105-126

Cizek, G. J. (2001). Conjectures on the rise and call of standard setting: An introduction to context and practice.

Setting performance standards: Concepts, methods, and perspectives (sayfa 3–17). Hillsdale, NJ: Erlbaum.

Cizek, G. J. ve Bunch, M. B. (2006) *Standard Setting: A Guide to Establishing and Evaluating Performance Standards on Tests*. Sage Publications.

Crocker, L. ve Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart and Winston, Inc.

Güler, N. (2008). *Klasik Test Kuramı Genellenabilirlik Kuramı ve Rasch Modeli Üzerine Bir Araştırma*.

Yayınlanmamış doktora tezi. Hacettepe Üniversitesi: Ankara.

Hauger, J. B. (2007). *An Empirical Comparison of the Bookmark and Modified Angoff Standard Setting Methods and the Impact on Student Classification*. Yayınlanmamış doktora tezi. University of Massachusetts: USA.

Irwin, P. (2007) *An Alternative Examinee-Centered Standard Setting Strategy*. Yayınlanmamış doktora tezi. University of Nebraska: USA.

Jeager, R. (1989). *Certification of Student Competence in Educational Measurement*. New York: American Council on Education and Macmillan, 485-514

Lee, G. ve Lewis, D. M. (2008). A Generalizability Theory Approach to Standard Error Estimates for bookmark standard settings. *Educational and Psychological Measurement*, 68(4), 603-620.

Mushquash, C. ve O'Connor, B. P. (2006). SPSS and SAS Programs for Generalizability Theory Analysis. *Behavior Research Methods*. 38 (3), 542-547.

Pallant, J. (2007). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using SPSS for Windows*. Open University Press.

Shavelson, R. J. ve Webb, N. M. (1991). *Generalizability Theory: A Primer*. Sage Publications, USA.

Sudweeks, R. R., Reeve, S. ve Bradshaw, W. S. (2005). A Comparison of Generalizability Theory and Many Facet Measurement in An Anlysis of College Sophomore Writing. *Assessing Writing*. 9, 239-261.