

Angoff, Ebel ve Nedelsky Yöntemleriyle Belirlenen Kesme Puanlarının Sınıflama Tutarlılıklarının Karşılaştırılması

Seçil ÖMÜR*, Hüseyin SELVİ**

Mersin Üniversitesi

Özet

Bu araştırmada Angoff, Ebel ve Nedelsky yöntemleriyle ve farklı yargıcı sayıları kullanılarak elde edilen kesme puanlarının sınıflama tutarlılığı indeksleri incelenmiştir. Araştırmanın çalışma grubunda bulunan yargıcıları; 2009-2010 eğitim öğretim yılında çeşitli illerin okullarında görev yapan 24 İngilizce öğretmeni ve Mersin Üniversitesi'nde görev yapan 6 alan uzmanı, öğrencileri ise; Mersin ilindeki özel bir dershaneye devam eden 176 ilköğretim 8. sınıf öğrencileri oluşturmaktadır. Yargıcı olarak nitelenen alan uzmanlarının ve öğretmenlerin puanlamaları sonucu yukarıda belirtilen yöntemlerle kesme puanları hesaplanmıştır. Araştırmanın sonunda, yargıcı sayısına bağlı olarak kesme puanlarında önemli bir değişim gözlenmemekle birlikte, üç yöntemde ve üç ayrı yargıcı sayısında en yüksek sınıflama tutarlılığı indeksinin Nedelsky yöntemiyle elde edildiği görülmektedir.

Anahtar Sözcükler: kesme puanı, Angoff yöntemi, Ebel yöntemi, Nedelsky yöntemi, sınıflama tutarlılığı

Abstract

In this study, cutoff scores' classification consistency indices calculated by Angoff, Ebel and Nedelsky methods and different rater numbers were examined. The participants of the study were 30 raters; 6 expert in the field working at Mersin university and 24 English teachers working at different government schools in 2009-2010 academic year; and 176 students, 8th grade secondary school students studying at private courses in Mersin. Cutting scores were calculated with the mentioned methods after the raters scored. The results of the study showed that there were not any significant difference in the cutting scores and also the highest classification consistency indice of three methods and three different rater numbers was observed via Nedelsky methods.

Keywords: cutting score, Angoff method, Ebel method, Nedelsky method, classification consistency

Ölçme araçlarından elde edilen puanlara dayanarak bireyler hakkında seçme, yerleştirme, tanı, rehberlik gibi çeşitli kurumsal kararlar verilmektedir. Geliştirilmiş bir ölçme aracını kullanarak bir birey için elde edilmiş bir puan hakkında karar verebilmek (assessment) için de çeşitli ölçütler kullanılmaktadır. Bu ölçütlerden biri de elde edilen puanın belirlenen bir kesme puanını aşmamasıdır. Kesme puanına dayalı değerlendirme işleminde teste giren bireylerin sınıflanması söz konusu olduğundan, verilecek olan kararlar da; geçer-kalır, başarılı-başarısız, yeterli-yetersiz, olumlu-olumsuz tutuma sahip vb. şeklinde genellikle iki kategori düzeyinde belirlenmektedir.

Jaeger (1989), standart belirleme yöntemlerini test merkezli veya öğrenci merkezli olarak ikiye ayırmıştır. Cizek (1996), Kane (1994) ve diğerleri bu ayrımı desteklemekte ve öğrenci merkezli yaklaşımlarda kesme puanının; bireylerin puan dağılımı üzerinden, çeşitli amaçlar doğrultusunda belirlenmekte olduğunu; test merkezli yaklaşımlarda ise; madde, güçlük düzeyi, çeldirici gibi değişkenlerden hareket edilerek kesme puanına ulaşıldığını belirtmektedirler.

Kesme puanının belirlenmesine ilişkin birçok farklı yöntem bulunmasına rağmen, her bir yöntemin uygulanmasında çeşitli sıkıntılar bulunmaktadır. Çünkü nasıl belirlenirse belirlensin, kesme puanını belirlemede bir keyfilik (arbitrary) söz konusudur. Genel olarak en iyi kesme puanı belirleme yönteminin bulunmadığı kabul edilmekle birlikte, her yöntemin kullanılabilirliğine ilişkin çeşitli

* Arş. Gör., Mersin Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı, secilomur@mynet.com

** Doktora öğrencisi, Mersin Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı, synapse333@hotmail.com

eleştiriler yapılmaktadır. Belirlenen kesme puanına dayalı olarak bireyler hakkında önemli kararlar verilmesi nedeniyle, bu kesme puanlarının belirlenmesinin yanı sıra, tutarlı sınıflama yapıp yapmadığına da bakılması gerekmektedir. Aşağıda bu çalışma kapsamında kesme puanı belirlemede kullanılan yöntemlerden bahsedilmektedir.

Angoff Yöntemi: 1971 yılında Angoff tarafından önerilen bu yöntemde, alanında uzman olan yargıcılardan, her bir madde için minimum yeterlikteki bireylerin o maddeyi doğru cevaplama yüzdesini belirlemeleri istenmektedir. Her madde için yargıcıların belirlediği bu yüzdelerin ortalamaları alınmaktadır. Her bir madde için ortalama geçme düzeyi belirlendikten sonra bu ortalamalar toplanmakta ve kesme puanı belirlenmektedir (Shultz & Whitney, 2004; Mills & Melican, 1998; Crocker&Algina, 1986).

Ebel Yöntemi: Çoktan seçmeli maddeler için uygun olan bu yöntemde, alanında uzman olan yargıcılara, ölçekteki bütün maddeleri güçlük (kolay, orta zor) ve önem derecelerine (gerekli, önemli, kabul edilebilir, sorgulanabilir) göre sınıflamaları için Tablo 1'deki gibi 3x4'lük bir tablo verilmektedir. Yargıcılardan maddelerin tamamını tabloda belirtilen hücrelere yerleştirmeleri ve minimum yeterlikteki bireylerin, sınıflanan bu maddelerin yüzde kaçını doğru cevaplayabileceğini belirtmeleri istenmektedir. Hesaplama aşamasında, her bir hücredeki madde sayısı ile o hücre için belirtilen geçme oranı çarpılmakta ve bu çarpımlar toplanmaktadır. Her bir yargıcı için bu işlem tekrarlanmakta ve ortalama alınarak nihai kesme puanına ulaşılmaktadır (Shultz & Whitney, 2004; Mills & Melican, 1998; Crocker&Alginia, 1986).

Tablo1: Ebel Yöntemiyle Maddelerin Güçlük ve Önem Derecelerinin Belirlenmesi

	Kolay	Orta	Zor
Gerekli Yüzdelerik (%)			
Önemli Yüzdelerik(%)			
Kabul Edilebilir Yüzdelerik(%)			
Sorgulanabilir Yüzdelerik(%)			

Nedelsky Yöntemi: Angoff yönteminden farklı olarak bu yöntemde maddelerle birlikte seçenekler de ele alınmaktadır. Bu nedenle sadece çoktan seçmeli maddeler için kullanılabilen bir yöntemdir. Bu yöntemde, alanında uzman olan yargıcılardan ölçekteki her bir maddenin çeldiricilerini incelemeleri ve minimum yeterlikteki bireylerin bu çeldiricilerden kaç tanesini eleyebileceklerini belirlemeleri istenmektedir. Daha sonra aşağıdaki formül uygulanarak her bir yargıcı için kesme puanı hesaplanır.

$$K=1/(\text{Ç.S.} - \text{E.Ç.S.})$$

K: Bir Yargıcının belirttiği kesme puanı

Ç.S.: Çeldirici sayısı

E.Ç.S.: Geçme-kalma sınırındaki bireyin eleyebileceği çeldirici sayısı

Her bir yargıcı için K'lar hesaplandıktan sonra yargıcılar arası ortalama alınarak nihai kesme puanına ulaşılr (Violato, Marini ve Lee, 2003; 64).

Yukarıda belirtilen kesme puanı belirleme yöntemlerinde yeterli yargıcı sayısının belirlenmesi, yargıcıların niteliklerine karar verilmesi, yargıcıların seçilmesi gibi sıkıntılar da bulunmaktadır. Ayrıca belirtilen farklı yöntemlerle belirlenen kesme puanları temel alınarak, bu puanların ne derecede tutarlı sınıflama yaptığının belirlenmesi de oldukça önemlidir. Bu yöntemleri karşılaştıran çeşitli çalışmalar bulunmasına rağmen (Chang, 1999; Livingston ve Zieky, 1989; Alsmadi, 2007), hiçbirinin sınıflama tutarlılığını incelemeye değeri görülmemiştir. Bu nedenle, yukarıda belirtilen farklı yöntemlerden elde edilen kesme puanlarından hangisinin, değişik yargıcı sayılarında daha tutarlı sınıflama yaptığının incelenmesi bu çalışmanın ana amacını oluşturmaktadır.

Araştırma Problemleri

1.a. Angoff, Ebel ve Nedelsky yöntemleriyle elde edilen kesme puanlarına dayanarak hesaplanan sınıflama tutarlığı indeksleri hangi yöntemde daha yüksektir?

1.b. Farklı yargıcı sayılarına (10, 20 ve 30) göre Angoff, Ebel ve Nedelsky yöntemleriyle hesaplanan kesme puanlarının, sınıflama tutarlığı indeksleri hangi yöntemde daha yüksektir?

Yöntem

Bu araştırma; Angoff, Ebel ve Nedelsky yöntemleriyle ve farklı yargıcı sayıları kullanılarak elde edilen kesme puanlarının sınıflama tutarlığı indekslerini incelediğinden temel bir araştırmadır.

Çalışma Grubu

Araştırmanın çalışma grubunda bulunan yargıcıları; 2009–2010 eğitim öğretim yılında çeşitli illerin ilköğretim okullarında görev yapan 24 İngilizce öğretmeni ve Mersin Üniversitesi'nde görev yapan 6 alan uzmanı, öğrencileri ise Mersin ilindeki özel bir dershaneye devam eden 176 ilköğretim 8. sınıf öğrencisi oluşturmaktadır.

İşlem

A.1. Yargıcılara, 8. sınıf düzeyinde 17 tane çoktan seçmeli test maddesinden oluşan ve bir dershanenin deneme sınavında kullanılan İngilizce testi verilmiştir (KR-20: .60). Yargıcılardan bu testin maddeleri için kendilerine verilen değerlendirme formları üzerinde,

- Angoff yöntemi için, geçme-kalma (minimum yeterlik) sınırındaki öğrencilerin o maddeyi doğru cevaplama yüzdesini tespit etmeleri,
- Ebel yöntemi için, ölçekteki bütün maddeleri güçlük ve önem derecelerine göre sınıflamaları ve geçme-kalma (minimum yeterlik) sınırındaki öğrencilerin sınıflanan bu maddelerin yüzde kaçını geçebileceklerini belirtmeleri ve
- Nedelsky yöntemi için, geçme-kalma (minimum yeterlik) sınırındaki öğrencilerin, her bir madde için eleyebileceği çeldirici sayısını belirtmeleri istenmiştir.

A.2. Farklı yargıcı sayılarına (10, 20 ve 30) göre Angoff, Ebel ve Nedelsky yöntemleriyle elde edilen kesme puanlarının sınıflama tutarlığı indeksleri karşılaştırılmıştır.

Verilerin Analizi

Burada; Angoff, Ebel ve Nedelsky yöntemlerine ilişkin kesme puanları farklı yargıcı sayıları da dikkate alınarak ayrı ayrı hesaplanmıştır. Farklı yargıcı sayıları (10, 20 ve 30) tüm yargıcılar içinden seçkisiz (random) seçilerek elde edilmiştir. Angoff, Ebel ve Nedelsky yöntemleri kullanılarak elde edilen kesme puanlarının sınıflama tutarlığı indeksleri, tek uygulamaya dayanan ve Subkoviak (1988)

tarafından geliştirilen ($|z| = \frac{(c - .5 - \bar{X})}{S}$) formülü kullanılarak hesaplanmıştır. Burada c=Kesme

Puanı, \bar{X} =Testin Ortalaması ve S=Testin Standart Sapması olarak belirtilmektedir. Subkoviak (1988) bu formül ile standart normal dağılımdan yararlanarak testin ortalaması, standart sapması ve KR-20 ya

da KR-21 kullanarak Po ve κ 'nın yaklaşık değerlerinin hesaplanabileceği tablolar geliştirmiştir (Akt: Erkuş, 1999; 50).

Bulgular

Farklı yargıcı sayısına göre; Angoff, Ebel ve Nedelsky yöntemleriyle elde edilen kesme puanları ve bu puanların sınıflama tutarlığı indeksleri aşağıda belirtilmektedir.

Tablo 2: Angoff, Ebel ve Nedelsky yöntemleri ile elde edilen kesme puanları ve bu puanların sınıflama tutarlığı indeksleri

Yargıcı Sayısı (N)	Angoff			Ebel			Nedelsky		
	Standart Sapma	Kesme Puanı	Sınıflama Tutarlılığı İndeksi (Po)	Standart Sapma	Kesme Puanı	Sınıflama Tutarlılığı İndeksi (Po)	Standart Sapma	Kesme Puanı	Sınıflama Tutarlılığı İndeksi (Po)
10	1.73	8.86	.72	1.60	8.98	.71	2.25	8.19	.74
20	1.05	8.88	.72	1.83	9.48	.70	1.62	8.42	.73
30	1.41	9.15	.71	1.71	9.12	.71	2.14	8.48	.73

Yukarıdaki tablo incelendiğinde, her bir yöntem ve farklı yargıcı sayıları için, ayrı ayrı değerlendirme yapıldığı görülmektedir. Bilindiği üzere Po bir orandır ve 0.00 ile 1.00 arasında değerler alır. Bu değer 1.00'a yaklaştıkça sınıflama kararlarının tutarlı olduğunu 0.00'a yaklaştıkça ise tutarsız olduğunu göstermektedir. Ayrıca; güvenilirlik arttıkça Po değerinin de artacağı ve bu değer .85 ve üzerinde olması gerektiği ileri sürülmektedir (Subkoviak, 1988; akt: Erkuş, 1999). Yargıcı sayısının 10 ve 20 olması durumlarında, en yüksek kesme puanının Ebel yöntemiyle ve en iyi sınıflamanın Nedelsky yöntemiyle elde edildiği görülmektedir. Yargıcı sayısının 30 olması durumunda ise, en yüksek kesme puanının Angoff yöntemiyle ve en iyi sınıflamanın Nedelsky yöntemiyle elde edildiği görülmektedir. Ayrıca yapılan bu değerlendirmeler incelendiğinde, her bir yöntem ve farklı yargıcı sayıları için standart sapma değerinin en yüksek Nedelsky yöntemiyle elde edildiği görülmektedir. Yargıcı sayısına bağlı olarak kesme puanlarında önemli bir değişim gözlenmemekle birlikte, üç yöntemde ve üç ayrı yargıcı sayısında en yüksek sınıflama tutarlığı indeksinin Nedelsky yöntemiyle elde edildiği görülmektedir.

Tartışma ve Yorum

Eğitimde ve psikolojide bireyler hakkında çeşitli kararlar verilirken, bireylerin ölçme aracından elde edilen ham puanları ölçüt olarak bir kesme puanı ile karşılaştırılabilmekte ve bireyler buna bağlı olarak sınıflanabilmektedir. Bireyler hakkında verilecek olan kararların bazen hayati öneme sahip olduğu düşünüldüğünde ise, sınıflama kararlarının ne derece isabetli olduğunun belirlenmesi büyük önem taşır. Ancak Angoff, Ebel, Nedelsky gibi yöntemlerle kesme puanı belirlenmesinde yargıcıların kullanılması sürece bir keyfilik karıştırmaktadır. Bu nedenle, yukarıda belirtilen yöntemlerle elde edilen kesme puanlarının hangisinin daha tutarlı sınıflama yaptığının incelenmesi gerekmektedir. Yargıcı sayıları ayrı ayrı dikkate alınarak Tablo 2'deki Po değerleri tekrar incelendiğinde ise; Nedelsky yöntemi kullanılarak hesaplanan kesme puanının, diğer yöntemlere göre daha tutarlı sınıflama yaptığı bu değer 10 kişilik bir yargıcı sayısından elde edildiği görülmektedir. Ayrıca; Po değerlerinin güvenilirliği yüksek ölçeklerde daha iyi sonuçlar vereceği düşünüldüğünden benzer çalışmaların, farklı güvenilirlik katsayılarına sahip ölçekler için elde edilen kesme puanlarının sınıflama tutarlılıkları karşılaştırılarak da yapılması gerektiği düşünülmektedir.

Kaynaklar

- Alsmadi, A. A. (2007). A comparative study of two standard-setting techniques. *Social Behavior And Personality*, 35(4) 479-486
- Chang, L. (1999). Judgmental item analysis of the Nedelsky and Angoff standard-setting methods. *Applied Measurement In Education*, 12(2) 151-165
- Crocker, L. ve Algina, J. (1986). *Introduction to classical and modern test theory*. Holt, Rinehart and Winston, Inc.
- Cizek, G. J. (1996). Standard- setting: guidelines, *Educational Measuremen:Issues And Practice*. 15
- Erkuş, A. (1999). *Ölçme araçlarının tutarlı ölçme ve sınıflama yapıp yapmadığını belirlemeye ilişkin bir çalışma*. Yayınlanmamış Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara,
- Jeager, R. (1989). *Certification of student competence in educational measurement*. New York: American Council on Education and Macmillan, 485-514
- Kane, M. (1994). Validating the performance standards associated with passing scores. *Review of Educational Research* 64 425-461.
- Livingston S. A ve Zieky M. J. (1989). A comparative study of standard-setting methods. *Applied Measurement In Education*, 2(2) 121-141
- Mills, C. ve Melican, G. (1998) Estimation and adjusting cut off scores: Features of selected methods, *Applied Measurement In Education*, I(3) 261-275 Lawrence Erlbaum Associates, Princeton.
- Shultz K. S. ve Whitney D. J. (2004) *Measurement theory in action: case studies and exercises*. Sage publications, Inc. 216,
- Violato, V., Marini, A. ve Lee C. (2003). A validity study of expert judgment procedures for setting cutoff scores on high-stakes credentialing examinations using cluster analysis, *Evaluation & The Health Professions*, Vol. 26, No. 1, Canada.