



Açık ve uzaktan öğretimde ölçme ve değerlendirme politikaları

Yrd. Doç. Dr. Murat AKYILDIZ^a

^a Anadolu Üniversitesi, Açıköğretim Fakültesi, Eskişehir, Türkiye 26470

Özet

Açık ve uzaktan öğretim çeşitli yönleriyle örgün öğretimden ayrılır. Bireylerin başarı durumlarının belirlenmesi amacıyla yapılan ölçme ve değerlendirme işlemlerinde üç ana politika bulunmaktadır. Bu çalışmada, ölçüt (kriter) referanslı, norm referanslı ve ipsative olmak üzere bu üç politikanın açık ve uzaktan öğretimde avantaj ve dezavantajları tartışılmış ve önerilerle bulunulmuştur.

Anahtar Sözcükler: Ölçme ve Değerlendirme, Standart Belirleme, Değerlendirme Politikaları, Açık ve Uzaktan Öğretim

Abstract

Open and distance education differentiates from formative education with same aspects. There are three main policies while assessing achievement of people. In this paper, advantages and disadvantages of criterion referenced, norm referenced and ipsative assessment methods in open and distance education have been discussed and some suggestions have been made.

Keywords: Measurement and Evaluation, Setting Standards, Assessment Policies, Open and Distance Education

Giriş

Açık ve uzaktan öğretim çeşitli yönleriyle örgün öğretimden ayrılır. Programın geliştirilmesi, içeriğin hazırlanması, içeriğin ulaştırılması gibi noktalara ek olarak ölçme ve değerlendirme süreçleri de örgün öğretimden farklılık gösterir. Ölçme ve değerlendirme süreçlerinde ortaya çıkan bu farklılıklar kendisini öncelikle sınav uygulaması sırasında gösterir. Aynı anda çok sayıda katılımcıya aynı ya da farklı içerikte bir sınavı ulaştırmak ve bu sınavın belirli standartlarda uygulanmasını sağlamak bugün geniş ölçekli sınav olarak adlandırılmaktadır (Bennett, 1998). Geniş ölçekli sınavlar hazırlanmaları, güvenliklerinin sağlanması ve değerlendirilme biçimleri bakımından örgün öğretimde uygulanan sınavlardan farklılıklar gösterirler. Geniş ölçekli sınavlar öncelikle çok sayıda kişinin hayatı üzerinde değişiklik yapma potansiyeli taşıyan sınavlardır. Bu anlamıyla da geniş ölçekli sınavlar politik bir yük taşımaktadır.

Bireylerin başarı durumlarının belirlenmesi amacıyla yapılan ölçme ve değerlendirme işlemlerinde üç ana politika bulunmaktadır. Bunlardan birincisi sınava girenlerin başarılarını bir ölçüte göre değerlendirmek (ölçüt referanslı değerlendirme), ikincisi başarıları başka katılımcıların başarılarına göre değerlendirmek (norm referanslı değerlendirme), üçüncüsü ise bireyleri eğitim programına başladıkları andaki başarı durumlarından ne kadar ileriye gidebildikleri bakımından değerlendirmektir (ipsative değerlendirme) (Hughes G., 2014; Hughes G., 2014; Hughes G., 2014; Hughes G., 2014; Hughes, Okumoto ve Crawford, 2010). Bu ana politikalara ek olarak karma bir politikadan da bahsedilebilir. Ülkemizde açık ve uzaktan öğretimde bir politika olarak ölçme ve değerlendirme süreçleri programların içeriklerinin belirlenmesine oranla göreceli olarak daha az gelişmiştir (Karadağ, 2014). Aşağıda tüm bu değerlendirme politikaları ve bu politikaların çeşitli durumlarda avantaj ve dezavantajları sunulmuştur. Bu çalışma ile özellikle açık ve uzaktan öğretim programlarında kullanılacak ölçme ve değerlendirme politikalarının belirlenme süreçleri ve bu politikaların avantaj ve dezavantajları Türkçe alanyazında ayrıntılarıyla ilk kez sunulmuş olacaktır.

Ölçüt Referanslı Değerlendirme Politikası

Bu değerlendirme sınava katılanların bir puana (kesme puanı) dayalı olarak başarılı olanlar ve olmayanlar şeklinde iki gruba bölüldüğü değerlendirme politikasıdır. Bu politika, katılımcıların belirlenen puandan itibaren dersin hedeflerini yeterince kazandığını kabul eder. Bu politikada dersi geçme notu hedeflerin kazanılma miktarı ile yüksek ilişki içindedir. Örneğin, geçme notu olarak 50 ya da 60 gibi bir puanı belirlemek ölçüt referanslı bir değerlendirme yaklaşımıdır. Bu

yaklaşımın temel sorusu şudur: Geçme notu kaç olmalıdır? Şunu bilmek gerekir ki geçme notunun hangi durumda kaç olması gerektiğine dair objektif ve genel geçer bir ölçü yoktur. Ayrıca geçme notunun kaç olması gerektiğine dair bir istatistiksel formül de bulunmamaktadır. Bir politika olarak bakıldığında bu yaklaşımla kesme puanının ne olacağından çok onu kimin belirlediği daha önemlidir. Ölçüt referanslı değerlendirme yaklaşımında kesme puanını iki odak belirleyebilir. Bunlardan birincisi eğitim sistemini yürüten politika koyucudur. Kurum ya da eğitim sisteminin bütünündeki yaygın eğitim politikaları doğrultusunda diğer politikalarla uyumlu olması bakımından uygun bir kesme puanı belirleyebilir. İkinci belirleyici ise bir uzman grubudur. Bu uzman grubu katılımcıların hangi becerilere sahip olma durumunda dersi geçmiş sayılacağına uzlaşarak karar verirler. Aşağıda uzman grubunun kesme puanı belirleme süreci ayrıntılarıyla anlatılmıştır.

Bir eğitim programından başarılı sayılacakların belirlenmesi için ölçüt belirleme süreci üç bileşenin bir araya gelmesini gerektirir. Bunlar sırasıyla moderatör, konu uzmanları ve politika koyucudur (Pitoniak ve Morgan, 2012).

Moderatör: Standart belirleme sürecini bilen ve konu uzmanlarının standart belirleme sürecini kurallara uygun yürütmesinde yol gösterici olacak olan kişidir. Bu kişinin standardı belirleme konusunda doğrudan bir etkisi olmamalıdır, böylece konu alanı uzmanları sonradan kararlarının değiştirilmesi tehlikesi hissetmeden bağımsız bir şekilde çalışabilmelidir. Moderatör kurum içinden olabildiği gibi kurum dışından da seçilebilir. Kurum içinden seçilen moderatörün uzman grubu için yanlılık oluşturabilecek özelliklere sahip olmaması beklenir. Kurum dışından gelen moderatörün ise kurum kültürü hakkında bilgilendirilmesi beklenir. Her iki durumda da moderatörün standart belirleme süreçleri hakkında eğitim almış olması temel ölçütdür. Moderatörün üç temel rolü vardır: 1) politika belirleyiciye uygun standardın seçilmesi konusunda yardımcı olmak ve araştırmayı desenlemek ve 2) standart belirleme çalışmasında uygun yöntemlerin kullanılmasını sağlamak ve uygun sonuçların elde edildiğinden emin olmak 3) Sürecin iyi bir standart belirleme süreci olabilmesi için tüm yöntemlerin olması gerektiği gibi işletildiğinden emin olmaktır. Bu görevlerin yerine getirebilmesi için moderatör şu işlemleri de yapmak zorundadır: 1) konu alanı uzmanlarını standart belirleme konusunda eğitmek, 2) konu alanı uzmanlarının tartışmalarını izlemek ve notlar almak ve 3) konu alanı uzmanlarına ihtiyaçları olan ek bilgileri sağlamak. Moderatörün kendi kişisel görüşlerini açıklaması ya da kendi kişisel görüşlerine uygun yönlendirmeler yapması standart belirleme sürecine zarar verir.

Konu alanı uzmanları: Standardı belirlenecek olan dersin uzmanları bu grubu oluşturmaktadır. Bu grubun temel görevi hangi yeterlilik düzeyinin dersi geçmek için yeterli olduğunu belirlemektir. Bu grubun üyeleri ilgili dersin verildiği fakültenin ya da programın öğretim elemanlarıdır. Grubun üyeleri programın öğretim elemanlarının iyi bir örnekleme olmalıdır. Böylece, hem dersin hem de programın yeterlilikleri birlikte garanti altına alınmış olunur. Ayrıca belirlenecek olan standardın daha geçerli olması sağlanmış olur.

Politika belirleyici: Konu alanı uzmanlarının ve moderatörün çalışmaları sonucu belirlenecek olan kesme puanının kendisine bildirileceği kişi ya da grup politika belirleyicidir. Politika belirleyici kendisine gelen kesme puanıyla ilgili olarak çeşitli değişiklikler yapabilir. Kurumun genel politikasına uymayan kesme puanları üzerinde düzeltmelere gidebilir.

Ölçüt Puanını Belirleme Süreci

Hangi puanının yeterli olan ve olmayan öğrencileri ayırdığını bulma işlemi aynı zamanda sınır öğrenciyi de tespit etme işlemidir. Bu nedenle kesme puanını bulabilmek için sınava girmiş öğrencilerin cevaplama yaptıkları kâğıtlar en yüksek nottan en düşük nota doğru sıralanır. Konu alanı uzmanları sınav içeriğinden yola çıkarak *sınırda* yeterliliğe sahip sınav kâğıdını (öğrenciyi) tespit etmeye çalışırlar. Bu aşamada bir kaç faktör, konu alanı uzmanlarının *sınırda* kâğıdı (öğrenciyi) belirlemelerine etki eder. Bu faktörlerin başında sınavda kullanılan madde türü gelmektedir. Çoktan seçmeli, uzun cevaplı ya da ikisinin birlikte kullanıldığı sınav durumlarından konu alanı uzmanlarının madde türlerini dikkate alarak değerlendirme yapması beklenir. Ayrıca aşağıda anlatılacak olan standart belirleme yöntemlerinin bazıları bazı madde türlerine uygun değildir. Ölçüt puanını belirlemek için kullanılan yöntemler iki grup altında incelenebilir.

1. Test Merkezli Yöntemler

- a. Angoff Yöntemi
- b. Doğrudan Uzlaşma Yöntemi
- c. İşaretleme (Bookmark) Yöntemi
- d. Evet/Hayır (Yes/No) Yöntemi
- e. Ebel Yöntemi
- f. Nedelsky Yöntemi

2. Sınava Katılanlar Merkezli Yöntemler

- a. Sınır Grup Yöntemi
- b. Zıt Gruplar Yöntemi
- c. İş Gövdesi (BoW, Body of Work) Yöntemi

Test merkezli yöntemler

Angoff Yöntemi: William Angoff'un 1971 yılında önerdiği bu yöntem (Angoff, 1971) her madde için maddenin *sınır* yeterlilikte birey tarafından doğru cevaplanma olasılığının yüzde olarak tahmin edilmesine ve ardından bu yüzdelerin toplanmasına dayanır. Örneğin, konu alanı uzmanları tarafından birinci maddenin *sınır* yeterlilikte birey tarafından doğru cevaplanabilme olasılığı 0.80 olarak belirlendiyse tüm maddeler için aynı işlem tekrarlanır ve bu yüzdeler toplanır. Elde edilen değer o sınav için bir geçme kalma *sınır* puanı olarak kabul edilir. Bazen düşünmeyi kolaylaştırabilmek için konu alanı uzmanlarına 100 *sınırdaki* yeterlilikte öğrenci olsa kaç bu soruyu doğru yapardı şeklinde sormak da aynı sonucu almaya yardımcı olmaktadır. Angoff yönteminin konu alanı uzmanları için zihinsel olarak yorucu olabildiğine dair bulgular da vardır (Humbleton, Brennan ve diğerleri, 2000); fakat Angoff yöntemi yaygınlıkla kullanılmaya devam etmektedir.

Doğrudan Uzlaşma Yöntemi: 2004 yılında Sireci, Hambleton ve Pitoniak tarafından önerilen bu yöntemde (Sireci, Hambleton ve Pitoniak, 2004), testi oluşturan maddeler ünite temelli olarak gruplara ayrılır. Örneğin, test 4 üniteden maddeler içeriyorsa 4 grup madde oluşturulur. Konu alanı uzmanlarına her bir grup madde verilerek *sınırdaki* yeterliliğe sahip öğrencinin bu maddelerin kaç tanesini doğru cevaplayabileceği sorulur. Uzmanların görüşlerinin ortalaması alınarak o grup sorular için geçme puanı belirlenir. Her grup için belirlenen ortalama puanlar toplanarak kesme puanı bulunmuş olur. Örneğin, her biri 5 sorudan oluşan 4 grup soru için sırasıyla 5,4,2 ve 1 madde *sınırdaki* yeterli öğrencinin yapabileceği soru sayıları olarak belirlenmiş olsun. Bu sınavın geçme puanı 20 üzerinden 12 olacaktır. Doğrudan uzlaşma yöntemi Angoff yöntemine göre daha kısa zamanda bitirilebilmektedir. Henüz çok yaygınlaşmamış bir yöntem olmakla birlikte doğrudan uzlaşma yönteminin maddeleri ünite bakımından alt gruplara ayrılabilir testler için uygun olduğunu söylemek gerekir.

İşaretleme (Bookmark) Yöntemi: Bu yöntemde testi oluşturan maddeler kolay olandan zora doğru dizilirler. Konu alanı uzmanları *sınırdaki* yeterliliğe sahip kişilerin hangi maddeden itibaren soruları doğru cevaplayamayacağını belirlemeye çalışır. Maddelerin güçlük indeksleri bu sıraya dizme işlemi için yardımcı olur. *Sınır* maddeyi belirleme işlemi bittiğinde o maddeye kadar olan puanların toplamı sınavın geçme puanını verir (Davis-Becker, Buckendahl ve Gerrow, 2011). Örneğin, her sorunun bir puan olduğu 20 maddelik bir sınavda *sınırdaki* yeterli kişi için belirlenen doğru yapabileceği son madde 10. madde ise o sınavın geçme puanı 20 üzerinden 10 olarak belirlenir. İşaretleme yöntemi, hızlıca *sınır* puan belirleme çalışmalarından sıklıkla kullanılmaktadır. Bu yöntem, madde güçlük indekslerinin önceden hesaplanmış olmasını gerektirmektedir. Testin içinde yer alan maddeler arası güçlük geçişlerinin büyük boşluklar içermeyecek şekilde düzenlenmiş olması bu yöntemin önemli gerekliliklerinden birisidir. Örneğin, 7. maddenin güçlüğü 0,50, 8. maddenin güçlüğü 0,20 olduğu bir durumda *sınırdaki* kişinin 7. maddeyi doğru yapabileceği kararı verilmişse birdenbire 8. Maddeyi doğru yapamayacağı kararı verilmek zorunda kalınır. Daha sonraki sınav uygulamalarında sınava güçlüğü 0,40 olan bir madde katıldığında *sınırdaki* kişinin bunu cevaplayabilip cevaplayamayacağı konu alanı uzmanlarında tartışılmış olmalıdır. Oysa yukarıdaki örnekte görüleceği gibi güçlüğü 0,40 olan bir madde testte bulunmadığından uzmanlarca tartışmaya dâhil edilmemiştir. Bu nedenle, işaretleme yönteminin tüm madde güçlüğü düzeylerinden soruları barındıran testlerde uygulanması daha doğru olacaktır.

Evet/Hayır (Yes/No) Yöntemi: Bu yöntem konu alanı uzmanlarına her maddeyi *sınırdaki* yeterliliğe sahip bir kişinin doğru cevaplayabilip cevaplayamayacağını sorulmasına dayanır. Konu uzmanları evet ya da hayır şeklinde cevap verir. Angoff yönteminin basitleştirilmiş hali olan bu yöntemde (Impara ve Plake, 1997) her madde için evet ya da hayır cevapları belirlenir. Sonuçta evet'lerin sayısı toplanarak test için geçme notu belirlenmiş olur. Örneğin, 20 maddelik bir sınavın 9 maddesi *sınırdaki* yeterli kişi için evet doğru yapabilir şeklinde kodlandıysa bu testin geçme notu 20 üzerinden 9 olarak belirlenmiş olur.

Ebel Yöntemi: 1972 yılında Ebel'in önerdiği bu yöntem (Ebel, 1972) maddeleri hem zorlukları bakımından hem de test için ne kadar gerekli oldukları bakımından sınıflandırmayı içerir. Bu sınıflandırma yapıldıktan sonra konu alanı uzmanları maddelerin *sınırdaki* yeterli kişi için doğru cevaplanma olasılıklarını belirlerler. Aşağıda Ebel'in önerdiği yönteme uygun bir tablo verilmiştir.

Tablo 1.			
<i>Ebel Yöntemi İçin Örnek Tablo</i>			
Gereklilik	Zorluk		
	Kolay	Orta	Zor
Olmazsa Olmaz			
Önemli			
Kabul Edilebilir			
Sorgulanabilir			

Yukarıdaki tabloda her konu alanı uzmanı sınavda yer alan maddeleri önce uygun hücreye yerleştirir. Örneğin, 1.,2.,5. ve 7. Maddeleri olmazsa olmaz orta zor madde görüyor olabilir. Bu maddeleri uygun hücreye yerleştirir. İkinci adımda, bu maddelerin *sınırdaki* yeterli kişi tarafından doğru cevaplanma ihtimalini yine aynı hücreye yazılır. Böylece, tüm maddeler gereklilik ve zorluk bakımından sınıflandırılmış olur. Üçüncü aşamada, her hücredeki doğru cevaplanma ihtimali yüzdesi ile soru sayısı çarpılır ve hepsi toplanır. Böylece, sınav için geçme notu elde edilmiş olur. Aşağıda örnek bir Ebel tablosu verilmiştir.

Tablo 2			
<i>Ebel Yöntemine Göre Doldurulmuş Örnek Bir Tablo</i>			
Gereklilik	Zorluk		
	Doğru oranı tahmini	Soru sayısı	Kategori için beklenen puan
Olmazsa Olmaz			
Kolay	95	3	$3 \times 0,95 = 2,85$
Orta	75	2	$2 \times 0,75 = 1,5$
Zor	50	1	$1 \times 0,50 = 0,5$
Önemli			
Kolay	80	3	$3 \times 0,80 = 2,4$
Orta	70	2	$2 \times 0,70 = 1,4$
Zor	50	1	$1 \times 0,50 = 0,5$
Kabul Edilebilir			
Kolay	90	2	$2 \times 0,90 = 1,8$
Orta	70	1	$1 \times 0,70 = 0,7$
zor	40	1	$1 \times 0,40 = 0,40$
Sorgulanabilir			
Kolay	60	2	$2 \times 0,60 = 1,2$
Orta	40	1	$1 \times 0,40 = 0,4$
zor	15	1	$1 \times 0,15 = 0,15$
TOPLAM			13,8

Görülebileceği gibi örnek sınavın geçme notu 20 soru üzerinden 13,8 olarak belirlenmiştir.

Nedelsky Yöntemi: 1954 yılında sadece çoktan seçmeli maddeler için Leo Nedelsky tarafından önerilmiş olan bu yöntemdir (Nedelsky, 1954) ve maddelerin seçeneklerini de kesme puanı belirleme sürecine dâhil etmektedir. Bu yöntemin temel süreci şu şekildedir: Konu alanı uzmanları her sorunun seçeneklerini incelerken *sınırdaki* yeterli bireyin kaç seçeneğin yanlış olduğunu bilebileceğini belirlemeye çalışırlar. Böylece, *sınırdaki* bireyin kaç seçenek arasında *gerçekten* düşünmek zorunda kalacağını tespit ederler. Son aşamada 1/eleyemediği seçenek sayısı formülü ile o maddenin puan değeri hesaplanır. Örneğin, konu alanı uzmanı *sınırdaki* yeterli kişinin, birinci soruda yer alan 5 seçeneğin 3'ünün yanlış olduğunu hemen tespit edebileceğini düşünebilir. Bu durumda bu sorunun puanı $\frac{1}{2}$ olacaktır. Başka bir soruda *sınırdaki* bireyin hiç bir seçeneği eleyemeyeceği tüm seçenekleri düşünmek zorunda kalacağı tespit edilmiş olsun. Bu durumda ise o maddenin puanı $\frac{1}{5}$ olacaktır. Tüm maddeler için puanlar toplandığında geçme notuna ulaşılabilecektir. Tüm uzmanların puanlarının ortalaması alınarak nihai geçme notu elde edilmiş olacaktır. Geniş ölçekli sınavlarda grubun yarısının geçmesinin yarısının kalmasının istendiği durumlarda Nedelsky yöntemi kullanışlı olmaktadır.

Sınava Katılanlar Merkezli Yöntemler

Sınır Grup Yöntemi: Livingston ve Zieky'nin 1982 yılında önerdiği bu yöntemde (Livingston ve Zieky, 1982), öğrencileri tanıyan bir grup konu alanı uzmanı belirlenir. Bu uzmanlar öğrencileri kesinlikle sınırın üzerinde, sınırdaki ve kesinlikle sınırın altında olmak üzere üç gruba ayırırlar. Sınırdaki olarak belirlenen grupta yer alan öğrencilerin puanları sıraya dizilir ve bu puanların medyanı alınır. Bu değer sınav için geçme puanı olarak belirlenir. Bu yöntem, öğrencilerin iyi tanınıyor olmasını gerektiren bir yöntemdir.

Zıt Grup Yöntemi: Yine Livingston ve Zieky tarafından 1982 yılında önerilmiş olan bu yöntem (Livingston ve Zieky, 1982), üç yerine iki grup kullanmayı içermektedir. Başarılı ve başarısız olarak belirlenen bu gruplar belirlenecek puan aralıklarına göre en yüksek puan en düşük puana doğru sıralanır ve her puan aralığındaki kişi sayılarının yüzdeleri çıkarılır. Daha sonra bu yüzdelere kümülatif olacak şekilde toplanır. Yüzdelere arası geçişlerin büyük olmasını engellemek amacıyla yüzdelere yumuşatma (smoothing) denen bir işlem uygulanır. Bu işlem, her bir puan grubunda başarılı ve toplam öğrenci sayılarının birbirine bölünmesiyle elde edilir. Yumuşatma işleminden sonra %50'ye denk gelen öğrencilerin puanı kesme puanı olarak belirlenir.

İş Gövdesi (BoW, Body of Work) Yöntemi: Bu yöntemin temel avantajı sınır grup yönteminde olduğu gibi öğrencilerin tanınıyor olmasını gerektirmemesidir (Kingston, Kahl, Sweeney ve Bay, 2001). Sınava giren öğrencilerin sınav kâğıtlarından seçilecek olan bir örneklem grubu iyi, sınır ve kötü olmak üzere üç gruba ayrılır. Daha sonra, sınır olarak belirlenen grubun kâğıtları yine okunarak iyi, orta ve kötü olmak üzere üç gruba ayrılır. Aynı işlem tek kâğıda ulaşıncaya kadar devam eder. Bu yöntem, kâğıtların defalarca okunmasını gerektiren bir yöntemdir.

Yukarıda ölçüt referanslı değerlendirme ve ölçüt belirleme süreçleri anlatılmıştır. Ölçüt referanslı değerlendirme çeşitli avantajlar ve dezavantajlar taşımaktadır (Esler ve Dziuban, 1974; Popham, 2014). En önemli avantajı, başarılı sayılan kişilerin kazandırılmak istenen becerilerin en azından alt sınırı kadar beceriye sahip hale geldiğinden emin olunabilmesidir. Bir diğer avantaj ise başarısızlık durumunda başarısızlığın hangi beceri alanından kaynaklandığını ayrıntılı olarak verebilmesidir.

Ölçüt referanslı değerlendirmenin bazı dezavantajları da bulunmaktadır. Bunların birincisi başarılı ya da başarısız öğrencilerin sayısını kestirme konusunda yeterli bilgi veremiyor oluşudur. Ölçütün yüksek tutulduğu durumda başarısız öğrencilerin sistemde yığılmasına yol açabilmektedir. Öte yandan, ölçüt çoğu kez değiştirilebilir olmadığından sistemde mezuniyet için bekleyen öğrenci sayısını denetim altında tutmanın tek yolu uygulanan sınavların zorluklarını değiştirmek olmaktadır. Böyle bir çözüm ise sınavların güvenilirlik ve geçerlilikleri üzerinde bozucu etkiye yol açabilir.

Norm Referanslı Değerlendirme Politikası

Norm referanslı değerlendirme, sınav uygulamasına katılan bireylerin başarılarını ya daha önce bu sınav uygulamasına katılanların başarılarıyla ya da şu anda sınava katılmakta olanların başarılarıyla kıyaslanması işlemidir. Bu nedenle iki tür norm referanslı değerlendirme vardır.

1. Normlara dayalı değerlendirme
2. Standart puanlara dayalı değerlendirme

Normlara Dayalı Değerlendirme

Norm, tipik olan, standartlaşmış olan demektir. Bu anlamıyla norm, eskiden beri bu halde olmakta olan anlamındadır. Fen Lisesinden mezun öğrencilerin üniversiteye giriş sınavında matematik bölümündeki doğru cevap sayılarının son 10 yıldaki ortalaması bir norm olarak düşünülebilir. Gelenekselleşmiş, öyle olmuş ve olmakta olan istatistiksel değerler ölçme ve

değerlendirme süreçlerinde norm olarak kabul edilirler. Açık ve uzaktan öğretim uygulamalarının ölçme ve değerlendirme süreçlerinde de normlar kullanılabilir. Normlar çeşitli değişkenler bakımından belirlenebilirler. Yaş, cinsiyet, mezun olunan lise, sınava girilen il ya da bu değişkenlerin bir kısmının kombine edilmiş halleri norm grubu belirlemek için kullanılabilir. Örneğin, 21 yaşında, meslek lisesinin okul öncesi öğretmenliği programından mezun olmuş kız öğrencilerinin matematik sınavının son 10 yıldaki ortalama net sayısı 20 soruda 7 ise bu puan, bu günden sonra sınava girecek 21 yaşında, meslek lisesinin okul öncesi öğretmenliği programından mezun kız öğrencileri değerlendirirken kullanılacak bir kesme puanı olacaktır. Basitçe söylemek gerekirse 7 net ve üzerinde doğru cevap vermiş olanlar matematik sınavından başarılı sayılırken 7 netin altında doğru cevap işaretleyebilmiş olanlar başarısız sayılacaklardır; fakat norm çalışması az önce söylendiği kadar basit bir çalışma değildir. Norm çalışmalarında kullanılan ortalamanın standart hatasının hesaplanarak yorumlara dâhil edilmesi beklenir. Yüksek standart hata seçilen örneklemin ortalamasının gerçek ortalamayı yansıtmama bakımından problem taşıdığını gösterir. Norm çalışmalarında düşük standart hata ile çalışılmak istenir. Standart hata, standart sapma/ \sqrt{n} formülü ile hesaplanır. Standart hatanın büyük ya da küçük olmasını etkileyen iki faktör vardır. Grubun puanlarının çeşitliliği (standart sapma) arttıkça (kişi sayısı aynı kalmak şartıyla) standart hata artmaktadır. Kişi sayısı arttığında ise (grubun puanlarının çeşitliliği aynı kalmak şartıyla) standart hata küçülmektedir. Bu nedenle, norm çalışmaları yürütülürken norm gruplarının kalabalık ve homojen gruptan oluşturulması beklenmektedir.

Alt grup sayısının çok fazla olduğu durumlarda norm çalışmalarının yürütülmesi zorlaşır. Her bir alt grup için norm üretmek ve bu normları kullanmak gerçekçi bir çözüm olmayabilir. Program sayısının ve her programda yer alan derslerin sayısının çok olduğu durumlarda her ders için norm üretilmeyebilir. Bu durumda norm yerine standart puanlar kullanılabilir. Standart puanlar aşağıda anlatılmıştır.

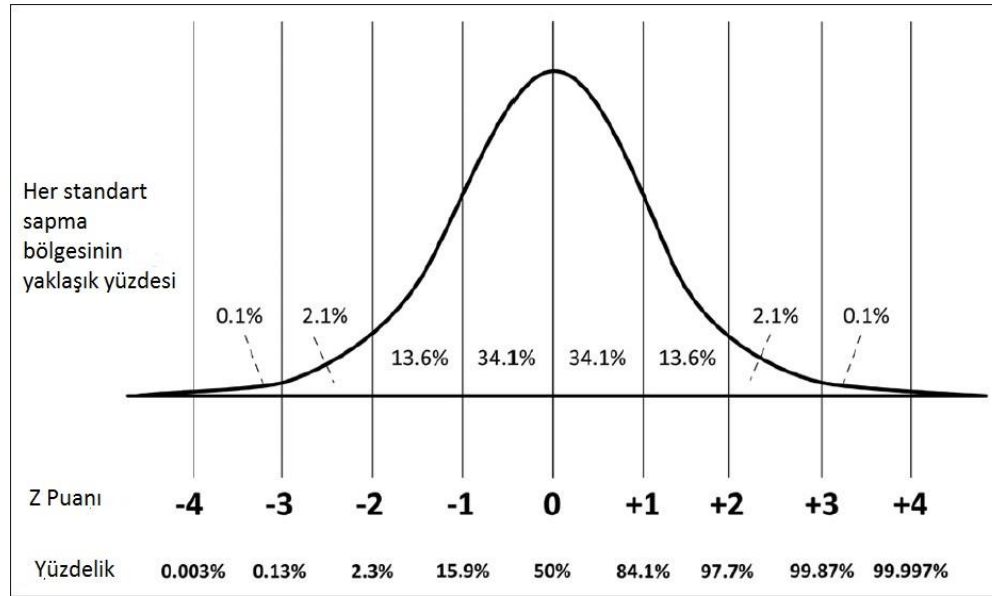
Norm çalışması yapıldığında puanlar ham halleriyle kullanılmazlar. Çeşitli norm puanları bulunmaktadır. Bunlar yüzdeler sırası ve normalize edilmiş z puanlarıdır

Yüzdeler Sırası: bir kişinin sırasının içinde bulunduğu gruptaki kişilerin yüzde kaçına denk geldiğinin bulunması yüzdeler sıranın bulunmasıdır. 100 kişilik bir grupta puanı bakımından 25. olan kişinin yüzdeler sırası %25'tir. Yüzdeler sırası sayesinde bir kişinin içinde bulunduğu gruptaki ya da norm grubundaki kişilerin ne kadarından daha başarılı ya da başarısız olduğu kararı verilebilir; fakat yüzdeler sıra'ya dayalı yorumların bazı dezavantajları vardır. Bunlardan birincisi kişi sayısının çoğaldığı puanlarda yüzdeler sırası güvenilirliğini yitirir. 100 kişilik bir grupta 25

kişinin 70, 25 kişinin 80 puan aldığı bir durum düşünün. İki gruptaki kişileri birbirinden ayırmak yüzdelerle sırası ile mümkün olmayacaktır. Ayrıca yüzdelerle sıralar doğrusal bir değişim göstermedikleri için iki yüzde arasındaki farka dayalı yorumlar istatistiksel olarak doğru olmamaktadır. Örneğin, 60 ve 65 puan arasındaki kişi sayısının oluşturduğu yüzdelerle fark %1 olabilirken 90 ve 91 puan arasındaki kişi sayısının oluşturduğu yüzdelerle fark %40 olabilir. Son olarak yüzdelerle sıralar eşit aralık ya da oran ölçeği düzeyinden bilgi vermedikleri için istatistiksel analiz yapılmasına da imkân tanımazlar. Örneğin, 60 puan alan %20 ile 65 puan alan %10'luk grubun ortalaması %15 şeklinde gösterilemez.

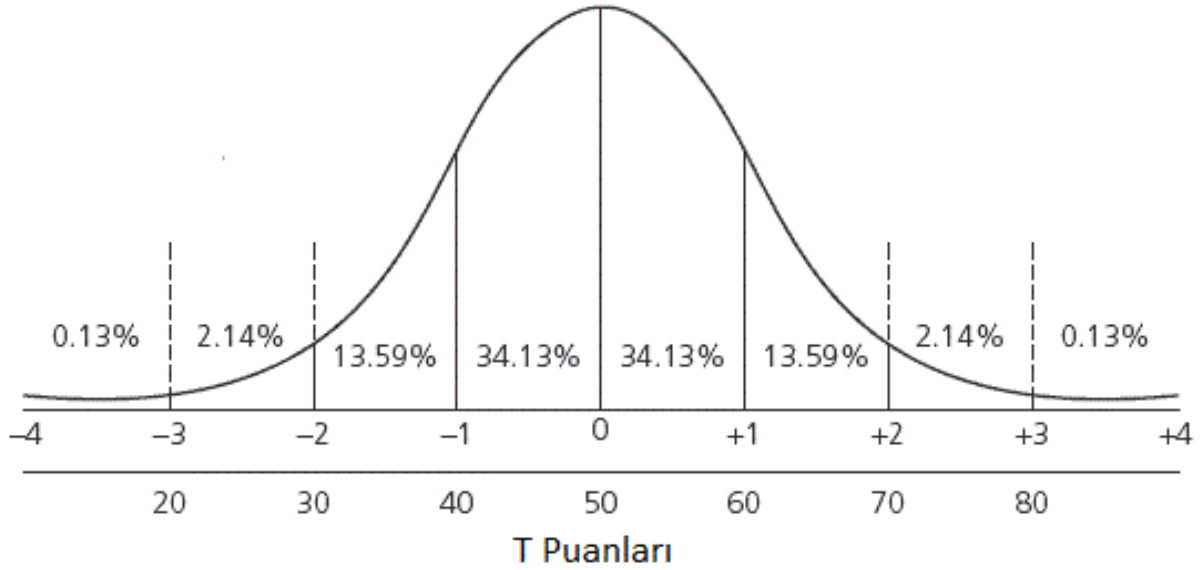
Yüzdelerle sıra kullanmak için en uygun olduğu durumlar öğrencilere içinde buldukları grupla kıyaslandığında başarılarının ne kadar olduğunu gösterilmesinin istendiği durumlardır. Ayrıca bir kesme puanı olarak düşünülecek olursa açık ve uzaktan öğretim yapan kurumun politikası her sene sadece belirli bir yüzdenin başarılı ya da başarısız olmasını garanti altına almak olduğunda yüzdelerle sıra kullanılabilir. Örneğin, programda kayıtlı bulunan öğrencilerin her sene sadece %25'inin mezun edilmesi planlanıyorsa geçme notu her sene puanların yüzdelerle sırasının %25'e denk geldiği yer olarak belirlenecektir.

Normalize Edilmiş Z puanları (Standart Puanlara Dayalı Değerlendirme): Z puanları uzun yıllardır ölçme alanyazınında kullanılmaktadır. Kısaca, Z puanları ham puanların ortalamadan uzaklıklarının standart sapma cinsinden ifadeleridir. Örneğin, bir kişinin puanı ortalamadan bir standart sapma uzakta ise kişinin Z puanı 1'dir. Normalize edilmiş Z puanları şu yöntemle elde edilir: Her puanın yüzdelerle değeri bulunur ve Z tablosundan o yüzdenin Z karşılığı elde edilir. Bu puanlar normal dağılıma çok yakın bir dağılım gösteren Z puanları olacaktır. Böylece, herhangi bir kişinin ortalamadan uzaklığına bağlı olarak başarısı hakkında karar verilebilir. Z puanlarının dağılımı aşağıda gösterilmiştir.



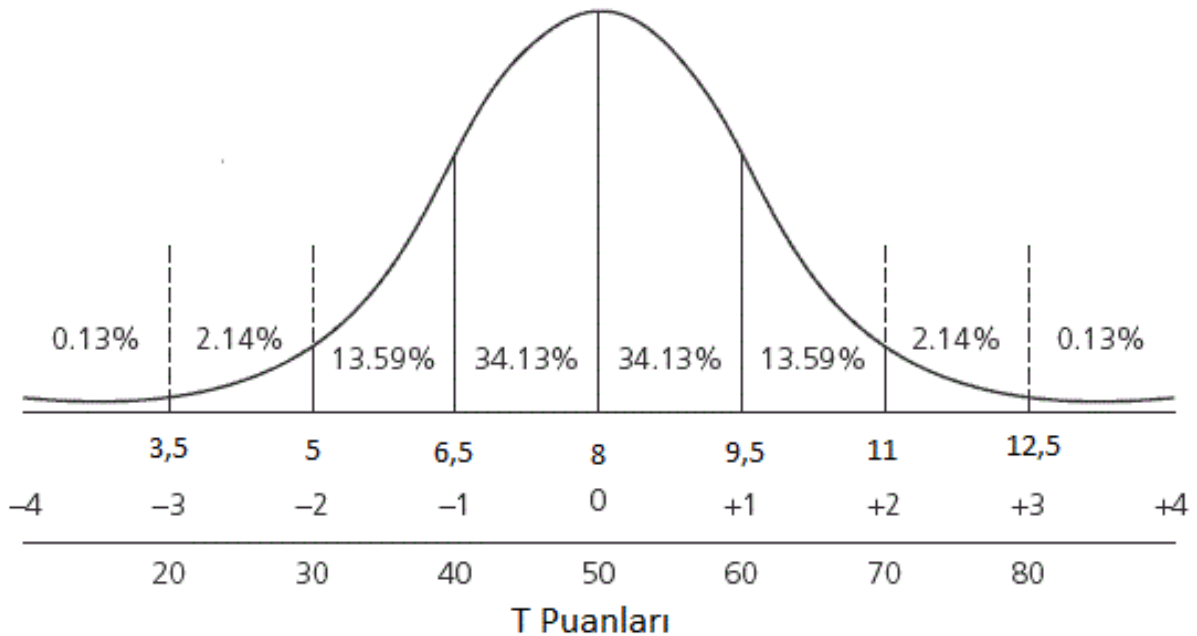
Grafik 1. Z puanlarının dağılımı

Yukarıdaki grafiğin basit bir yorumu şöyle yapılabilir. Z puanı +3 olan bir kişi içinde bulunduğu grubun yaklaşık %99,87'sinden daha başarılıdır. Benzer şekilde z puanı -2 olan bir kişi ise içinde bulunduğu grubun yaklaşık %2,3'ünden daha başarılıdır. Z puanı -1 le 0,00 arasında yer alan iki kişi arasında gruptaki bireylerin %34,1'i bulunmaktadır. Bir uzaktan ya da açık öğretim kurumu kendi üst politikaları doğrultusunda herhangi bir z puanını geçme notu olarak belirleyebilir. Hangi z puanının geçme notu olarak belirleneceğine dair akademik herhangi bir ölçüt ya da istatistiksel bir formül yoktur. Z puanlarına dayalı değerlendirmenin negatif değer alıyor olması nedeniyle çeşitli zorlukları olabilir. Z değerleri doğrusal bazı dönüştürmelerle tamamen pozitif değerler alacak hale dönüştürülebilir. Örneğin, Z puanlarının ortalamasının 50 standart sapmasının 10 olarak dönüştürülmesi durumunda T puanı denen bir puan türü üretilmiş olur. T puanları z puanları ile aynı anlama gelirler. Aşağıda Z puanlarına karşılık gelen T puanları gösterilmektedir.



Grafik 2. Z puanlarına Karşılık Gelen T Puanlarının Dağılımı

Görüleceği gibi T puanları Z puanlarına karşılık gelen ortalaması 50 standart sapması 10 olan puanlardır. Örneğin, Z puanı -1 olan bir kişinin T puanı 40 olmaktadır. T puanları gruba dayalı kıyaslamalarda sıklıkla kullanılmaktadır. Aşağıda 20 soruluk bir sınavdan alınmış hipotetik bir puan dağılımının Z ve T puanı karşılıkları verilmiştir. Bu puan dağılımının ortalamasının 8 standart sapmasının 1,5 olduğu varsayılmıştır.



Grafik 3. Örnek Bir Puan Dağılımının Z ve T Puanı Cinsinden Karşılıkları

Görüleceği gibi sınava katılanların ortalaması 8'dir. 8 puan almış öğrencilerin Z puanı 0 olurken T puanları ise 50'dir. Bu sınavdan 11 puan almış olan öğrencilerin Z puanları +2, T puanları ise 70'tir. Bu anlamıyla T puanları daha dar ya da daha geniş aralıktaki puanları ortalama 50 olacak şekilde 100'lük puan sistemi içine yaymakta ya da sıkıştırmaktadır. Zaman zaman ortalamanın 50 yerine 100 alındığı durumlar da olabilmektedir. Kişi sayısının fazla olduğu ya da değişkenliğin küçük olduğu durumlarda veya küçük farkların gösterilmesinin önemli olduğu durumlarda ortalama 100 standart sapma 25 olarak alınabilmektedir. Böylece, puanlar daha geniş bir aralığa yayılabilmektedir.

Norm dayanaklı değerlendirme politikasının bazı avantaj ve dezavantajları bulunmaktadır (Chan, 2014; Crocker ve Algina, 2006). Norm dayanaklı değerlendirme, sadece puana bakarak kişileri içinde buldukları gruptaki diğer kişilerden ne kadar iyi ya da kötü olduğu bakımından kıyaslamaya imkan tanır. Kısa sürede bireylerin başarı sıraları hakkında bilgi verir. Norm referanslı değerlendirme politikası (ortalama üstü değerlerin özellikle referans alındığı nadir durumlar hariç) her zaman grubun en az yarısının başarılı sayılabildiği bir platform sağlar. Ortalamanın altındaki değerlerin başarı standardı olarak alındığı durumlarda başarılı sayılan kişi sayısı grubun yarısından da fazla hale gelir. Ölçüt referanslı değerlendirme sisteminde başarılı ya da başarısız kişi sayısının ne kadar olacağını belirleyebilmenin mümkün olmadığı düşünüldüğünde bu imkân norm referanslı değerlendirmenin önemli bir avantajıdır. Norm referanslı değerlendirme politikası ile düşük başarılı kişi sayısının baskın olduğu kalabalık sistemlerde başarısızlık nedeniyle oluşan yıl tekrarları ve kişi sayısı artışları engellenerek kişi sayısından kaynaklanan enflasyonlar engellenmiş olur.

Yukarıda sayılan avantajlara rağmen bu politikanın bazı dezavantajları da bulunmaktadır. En büyük dezavantajı öğrenme hedeflerine ulaşma konusunda herhangi bir garanti sunmamasıdır. Programa katılanların programın hedeflerine ulaşmadan da programı başarılı olarak bitirebilme ihtimali bulunmaktadır. Norm referanslı değerlendirmenin bir diğer dezavantajı ise bireysel başarı ilerlemesi hakkında sınırlı bilgi veriyor olmasıdır. Farklı yarıyılarda iki sınavın birisinden sınıf ortalamasına eşit ikinci yarıyıl da sınıf ortalamasından yüksek puan alan bir kişinin başarısında ilerleme olduğunu söylemek mümkün olmayacaktır. Puanların içinde bulunulan gruba göre üretiliyor olması bireysel başarı ilerlemesini izlemeyi imkânsız hale getirmektedir.

İpsative Değerlendirme

Bir eğitim programına katılan kişilerin eğitim programının başlangıcından itibaren gösterdikleri ilerlemeyi bir başarı göstergesi olarak almak ipsative değerlendirme olarak bilinmektedir. İpsative değerlendirmede rekabet kişinin kendisiyledir. Yapılan çalışmalar özellikle uzaktan öğretimde başarı gösteremeyen öğrencilerin motive edilmesi bakımından önemli avantajlar getirdiğini göstermektedir (Hughes, Okumoto ve Crawford, 2010). Kişilerin programa başladıkları andan itibaren gösterdikleri ilerlemenin eğitimi alan birey tarafından görülebilir hale gelmesi kendi eğitim süreçlerini ve bu süreçlerin problem oluşturan noktalarını düzenleme bakımından katkılar getirmektedir (Hughes, 2014).

İpsative değerlendirme, not baskısından çok öğrenme miktarına dikkat çektiği için açık ve uzaktan öğretim hedefleriyle de daha uyumludur. İpsative değerlendirme yapılırken çık ve uzaktan öğretim programına kayıtlı kişilerin hem dönem içi hem de dönemler arası başarı değişimleri bildirilebilir. Böylece, hem dersler hem de dönem bazındaki başarı ilerlemesi görülebilir hale gelmiş olur. Türkiye’de açık ve uzaktan öğretim uygulamalarında ipsative değerlendirme yaygın değildir. Bu konuda çalışmalara ihtiyaç duyulmaktadır.

Karma değerlendirme politikası

Son yıllarda hem örgün hem de açık ve uzaktan öğretim programlarında bir kaç değerlendirme politikası birlikte kullanılmaktadır. Yaygın olan yöntem, geçme puanı olarak bir alt limit belirlemek (ölçüt referanslı değerlendirme) bu limitin üzerinde bulunan puanları ise norm referanslı değerlendirme ile ele almaktır. Örneğin, geçme puanı için alt limit 100 puan üzerinden 25 olarak belirlenip 25 üzerindeki puanlar norm referanslı değerlendirmeye tabi tutulmaktadır. 25 üzerinde alınan puanların en büyüğü AA harf notuna dönüştürülmektedir. Buna göre grubun başarı durumu dağılımına bağlı olarak bir sınavda 40 puan AA olabildiği gibi diğer bir sınavda 90 puan AA olabilmektedir. Karma değerlendirme politikası özellikle sistemin belirli miktarda öğrenmeyi garanti altına alıp bu öğrenme miktarının üzerindeki öğrenmeleri ödüllendirmek istediği durumlarda kullanılmaktadır.

Sonuçlar

Bir politika olarak ölçme ve değerlendirme üç temel politikayı içinde barındırmaktadır. Her politikanın açık ve uzaktan öğretim sistemine sağlayacağı avantajlar açık ve uzaktan öğretim sisteminin üst politikalarıyla doğrudan ilişkilidir. Ölçüt referanslı değerlendirme politikası tüm dünyada popülerliğini yitirmiş gibi görünmekle birlikte tamamen terk edilmiş

değildir. Unutmamak gerekir ki ölçüt referanslı değerlendirme için kesinleşmiş bir “iyi bir geçme notu” sınırı yoktur. Norm referanslı değerlendirme ise geniş katımlı sistemlerde öğrenme miktarını garanti altına alamaması nedeniyle dikkatle kullanılmalıdır.

İpsative değerlendirme ülkemizde ne örgün ne de açık ve uzaktan öğretimde kullanılmamıştır. Bu konuda yapılacak çalışmaların özellikle güdü ve kuruma bağlılığı arttıracığı düşünülmektedir. İpsative değerlendirmenin ayrı ayrı hem ölçüt referanslı hem de norm referanslı değerlendirme politikası ile birlikte kullanılabilir olması büyük bir avantajdır. İpsative değerlendirmenin ölçüt referanslı ve norm referanslı karma değerlendirme sistemlerine bütünleştirilmesi de mümkündür. Böylece kurumlar, hem belirli miktarda öğrenmeyi garanti altına alabilir, hem bu limit öğrenmenin üzerindeki çabayı ödüllendirebilir hem de bireylerin geçmiş başarılarıyla kıyaslamalar yaparak eksiklerin giderilmesi yönünde norm referanslı değerlendirmenin dezavantajlarını ortadan kaldırabilirler. Genel geçer, her zaman doğru kabul edilebilecek bir ölçme ve değerlendirme politikası yoktur. Tüm ölçme ve değerlendirme politikaları karma sistemler de dâhil olmak üzere avantaj ve dezavantajlar barındırmaktadır. Açık ve uzaktan öğretim kurumlarının politika belirleyicileri ilgili avantaj ve dezavantajları göz önünde bulundurarak bir ölçme ve değerlendirme politikasına karar vermelidir.

Öneriler

Açık ve uzaktan öğretimde ölçme ve değerlendirme politikasına karar verilirken kurumun üst politikası ile uyumlu bir politika belirlenmelidir. Açık ve uzaktan öğretim programlarının hedefleri ile uyumsuz bir ölçme ve değerlendirme sistemi sürekli güncellemeye ihtiyaç duyan bir yapı sergileyecektir. Bu sık güncelleme ihtiyacı kurumun enerjisinin gerçekte ihtiyaç duyulan alanlara yönelmesini engelleyecektir.

Ölçüt referanslı ve norm referanslı değerlendirme sistemlerini karşılaştıran çalışmalar olmasına rağmen ipsative değerlendirme üzerine yapılan çalışmalar yok denecek kadar azdır. Uygun açık ve uzaktan eğitim programlarında ipsative değerlendirmenin tek başına ya da karma bir sistem olarak uygulanmasının uygulamada getireceği avantaj ve dezavantajların incelenmesi alandaki önemli bir bilgi eksikliğini giderecektir.

Kaynakça

- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (pp. 508-600). Washington DC: American Council on Education.
- Bennett, R. E. (1998). *Speculations on the future of large-scale educational testing*. Princeton, NJ: Educational Testing Service.
- Chan, W. (2014). A better norm-referenced grading using the standard deviation criterion. *Teaching and Learning in Medicine*, 26(4), 364-365.
- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Belmont, CA: Cengage Learning.
- Davis-Becker, L. S., Buckendahl, C. W., & Gerrow, J. (2011). Evaluating the bookmark standard setting method: The impact of random item ordering. *International Journal of Testing*, 11, 24-37.
- Ebel, R. L. (1972). *Essentials of educational measurement*. Englewood Cliffs, NJ : Prentice Hall.
- Esler, W. K., & Dziuban, C. D. (1974). Criterion referenced test: Some advantages and disadvantages for science instruction. *Science Education*, 58(2), 171-174.
- Hughes, G. (2014). *Ipsative assessment motivation through marking progress*. Basingstoke, UK: Palgrave Macmillan.
- Hughes, G., Okumoto, K., & Crawford, M. (2010). *Centre for distance education teaching and research awards round 5 project Report*.
- Impara, J. C., & Plake, B. S. (1997). Standard setting: An alternative approach. *Journal of Educational Measurement*, 34, 353-366.
- Karadağ, N. (2014). Açık ve uzaktan eğitimde ölçme ve değerlendirme: Mega üniversitelerdeki uygulamalar. *Yayınlanmamış Doktora Tezi*. Eskişehir: Anadolu Üniversitesi.
- Kingston, N., Kahl, S., Sweeney, K., & Bay, L. (2001). Setting performance standards using the body of work method. In G. J. Cizek (Ed.), *Setting Performance Standards, Concepts, Methods, and Perspectives* (pp. 219-248). Mahwah, N.J: Erlbaum.
- Livingston, A., & Zieky, M. J. (1982). *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service.
- Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14, 3-19.
- Pitoniak, M. J., & Morgan, D. L. (2012). Setting and validating cut scores for tests. In C. Secolsky and D. B. Denison (Eds.), *Handbook on Measurement, Assessment, and Evaluation in Higher Education* (pp. 345-346). UK: Roudledge.
- Popham, W. J. (2014). Criterion-referenced measurement: Half a century wasted? *Educational Leadership*, 62-66.
- Sireci, S. G., Hambleton, R. K., & Pitoniak, M. J. (2004). Setting passing scores on licensure exams using direct consensus. *CLEAR Exam Review*, 15(1), 21-25.

Yazar Hakkında

Yrd. Doç. Dr. Murat AKYILDIZ



Psikoloji lisans eğitimi aldıktan sonra yüksek lisansını ve doktorasını ölçme ve değerlendirme alanında tamamladı. Çok boyutlu ve çok parametrelili lojistik modellerle bireylerin bilişsel ve duyuşsal özelliklerinin kestirilmesi temel ilgi alanıdır. Diğer ilgi alanları arasında psikometri, istatistiksel analiz, bilim yöntemi ve felsefesi yer almaktadır. Dr. Akyıldız, halen Anadolu Üniversitesi Açıköğretim Fakültesi öğretim üyesi olarak görev yapmaktadır.

Posta adresi: Anadolu Üniversitesi Açıköğretim Fakültesi
Tel (İş): +90 222 3350580/5627
Eposta: muratakyildiz@anadolu.edu.tr