

## Dereceli Puanlama Anahtarı ile Öz, Akran ve Öğretmen Değerlendirmesi

### Self-, Peer- and Teacher-Assessment Through Rubrics

Perihan GÜNEŞ<sup>1</sup>

Didem KILIÇ<sup>2</sup>

Başvuru Tarihi: 09.03.2016

Yayına Kabul Tarihi:23.09.2016

DOI: 10.21764/efd.93792

**Özet:** Son yıllarda performans değerlendirme için kullanılan araçlar arasında dereceli puanlama anahtarlarının sıklıkla kullanıldığı ayrıca öz değerlendirme ve akran değerlendirme uygulamalarının eğitim araştırmalarında önemli bir yer edindiği görülmektedir. Bu çalışma kapsamında öğretmen adaylarının sözlü sunum performansları dereceli puanlama anahtarı ile akranları, kendileri ve öğretmenleri tarafından değerlendirilmiş ve sonrasında öz, akran ve öğretmen değerlendirmeleri karşılaştırılmıştır. Öğretmen adaylarının performans değerlendirmeye ve dereceli puanlama anahtarı kullanımına ilişkin yeterliklerinin geliştirilmesine katkı sağlaması hedeflenen çalışmada var olan bir durum ortaya konulmaya çalışıldığından betimsel yöntem izlenmiştir. Araştırmanın çalışma grubunu, 18 öğretmen adayı ve 1 alan eğitimi uzmanı oluşturmuştur. Her bir öğretmen adayının performansı kendisi, akranları ve öğretmeni tarafından analitik türde geliştirilmiş dereceli puanlama anahtarı aracılığıyla değerlendirilmiştir. Öz, akran ve öğretmen değerlendirmelerinin karşılaştırılmasına yönelik araştırma probleminin çözümü için veriler Friedman testi ile analiz edilmiştir ve sonucunda değerlendirmeler arasında anlamlı farklılıklar olduğu tespit edilmiştir. Öğretmen adaylarının performanslarına ilişkin olarak yaptıkları öz değerlendirmelerin, akranları ve öğretmenleri tarafından yapılan değerlendirmelerden istatistiksel olarak anlamlı biçimde düşük olduğu belirlenmiştir. Akran değerlendirmesi ile öğretmen değerlendirmesi arasında bulunan farkın ise istatistiksel olarak anlamlı olmadığı sonucuna ulaşılmıştır.

**Abstract:** In recent years, rubrics have gained greater priority among performance evaluation tools to evaluate student performance. Besides, self- and peer-assessment applications have aroused great deal of interest in educational research. Within the scope of the study pre-service teachers' oral presentation skills were evaluated by themselves and also by their peers and a teacher through a rubric. The present study aims to compare self-, peer- and teacher-assessment of the pre-service teachers' performances through a rubric. The study group of the current research consists of 18 pre-service teachers and 1 subject area education expert. Performance of each pre-service teacher was evaluated by themselves, peers and teacher through the rubric developed in the analytic format. In order to compare the self-, peer- and teacher-assessment, Friedman test was run. Pre-service teachers' self-assessments of their oral presentation performances are found to be statistically significantly lower than peer- and teacher-assessments of their performances. When the scores obtained from peer- and teacher-assessments were compared, no significant difference was observed.

**Anahtar sözcükler:** *dereceli puanlama anahtarı, performans değerlendirme, öz değerlendirme, akran değerlendirme*

**Keywords:** *rubric, performance evaluation, self-assessment, peer-assessment*

<sup>1</sup> Aksaray Üniversitesi, Eğitim Fakültesi, İlköğretim Bölümü, Aksaray, Türkiye

<sup>2</sup> Aksaray Üniversitesi, Eğitim Fakültesi, İlköğretim Bölümü, Aksaray, Türkiye

## Giriş

Günümüzde izlenen yeni öğretim yaklaşımlarının öğrenme ortamlarına yansması sonucunda öğrencilerin bilimsel süreç becerilerinin geliştirilmesi, yaratıcı ve eleştiren, araştırmacı ve sorgulayıcı bireyler olarak yetiştirilmesi önem kazanmakta ve öğrencilerin gerçek yaşam durumlarına daha hazırlıklı olmaları teşvik edilmektedir. Tüm bu yeni eğilimler, eğitimin tüm bileşenlerine olduğu gibi, sınıf içi ölçme ve değerlendirme yaklaşımlarına da yansmıştır (Berberoğlu, 2006). Okullarda kullanılan ölçme değerlendirme yaklaşımları bilgilerin ne derece kazanıldığını ölçen bir yapıdan, bilgilerin yeni durumlarda ya da gerçek yaşam durumlarında nasıl kullanılabileceğini ölçen bir yapıya dönüşmüştür (Kutlu, Doğan, & Karakaya, 2008). Sınıf içinde sıklıkla kullanılan ölçme ve değerlendirme yaklaşımlarının (yazılı, sözlü, kısa yanıt, çoktan seçmeli vb sınavlar) yanı sıra performansa dayalı değerlendirme yaklaşımlarına yer verilmesi, bu yaklaşımların birlikte ve dengeli bir şekilde kullanılması, öğrencilerin öğrenmesine ve gelişimine ilişkin daha bütüncül ve doğru bilgi toplanabilmesi adına etkili sonuçlar verecektir. Öğrencilerin bilgiyi kavrayıp kavramadıkları hakkında tespit için mevcut ölçme ve değerlendirme araçları etkili olurken, sahip oldukları bilgiyi gerçek durumlarda ne oranda uygulayabilecekleri konusunda tespit için ise performansa dayalı değerlendirme araçları etkili olacaktır (MEB, 2007). Son yıllarda performans değerlendirme için kullanılan araçlar arasında dereceli puanlama anahtarlarının okullarda sıklıkla kullanıldığı ayrıca öz değerlendirme ve akran değerlendirme uygulamalarının eğitim araştırmalarında önemli bir yer edindiği görülmektedir.

Öz değerlendirme, öğrencilerin kendi öğrenmeleriyle ilgili kendilerini değerlendirmeleri anlamına gelirken (Boud & Falchikov, 1989, s. 529) akran değerlendirme, bir grup içerisinde yer alan bireylerin akranlarını değerlendirmeleri olarak tanımlanmaktadır (Falchikov, 1995, s. 176). Öz ve akran değerlendirme ile öğrenciler, hem kendilerinin hem diğer öğrencilerin performanslarının veya öğrenme ürünlerinin niteliği, derecesi ve değeri hakkında karar vererek değerlendirme sürecine katılırlar (Somervell, 1993; Topping, 2009). Öz değerlendirme, öğrencilerin kendi öğrenme süreçlerine aktif olarak katılmalarını sağlarken, kendi gelişimlerini yansıtmaya ve kendilerini eleştirmelerine olanak tanır (Dochy, Segers, & Sluijsmans, 1999). Öğrenciler akranlarını değerlendirirken de, kendi performanslarına ilişkin fikir sahibi olmaktadır (Bostock, 2001). Akran değerlendirme süreci ile farklı bakış açılarından geribildirim almak öğrencilerin akranları ile kendileri arasındaki farklılıklarını görmelerine ve kendilerine ilişkin farkındalık düzeylerini artırmalarına imkan tanımaktadır (Saito & Fujita, 2004). Bu durum öğrencilerin kendi öğrenmelerini düzenlemelerine böylece ilerideki öğrenmelerinin kolaylaşmasına yardımcı olmak gibi önemli bir fayda sağlamaktadır.

Öz değerlendirme ve akran değerlendirme birbirlerini destekleyen avantajlara sahip olduğundan öğrenme ortamlarında bu uygulamalara sıklıkla bir arada yer verilmektedir. Araştırmacılar öz değerlendirme ve akran değerlendirmenin, öğrencileri motive ettiği ve aktif katılımı sağladığı, etkili öğrenme gerçekleştirdiği, öğrencilerin öğrenmelerinin sorumluluğunu ve kontrolünü almalarına destek olduğu ayrıca öğrenmelerini daha iyi yapılandırabilmelerini sağlamak gibi güçlü yanlarının olduğunu ifade etmektedirler (Sluijsmans, Dochy, & Moerkerke, 1999). Öz ve akran değerlendirme süreçlerinin öğrencilerin, sorumluluk alma, muhakeme ve otonomi becerilerinin gelişmesine katkı sağladığı ifade edilmektedir (Sluijsmans ve diğ., 1999). Bu olumlu özelliklerinin yanı sıra öz ve akran değerlendirme uygulamalarının geçerlik ve güvenilirliği ile ilgili tartışmalı durumların önüne geçebilmek için uygulamaya yönelik çeşitli öneriler sunulmaktadır (Lindblom-Ylänne, Pihlajamäki, & Kotkas, 2006). Belirli ölçütlerin kullanılması, ölçme sürecinin şeffaf olması, yönergelerin açık ve anlaşılır olması, puanlama matrislerinin kullanılması gibi yaklaşımların bu uygulamalara katkı sağlayacağı savunulmaktadır. Öz değerlendirme ve akran değerlendirme için dereceli puanlama anahtarının kullanılmasının uygun olacağı belirtilmekte (Goodrich,

1997) ayrıca dereceli puanlama anahtarının, öğrenci performansını değerlendirmek için hem faydalı hem de ilgi çekici bir araç olduğu vurgulanmaktadır (Jensen, 1995).

Dereceli puanlama anahtarı, herhangi bir çalışma için dikkate alınan ölçütlerin listelendiği ve her bir ölçütün niteliğini iyiden kötüye doğru detaylı tanımlamalarla ortaya koyan doküman olarak tanımlanmaktadır (Andrade, 2000; 2001; 2005; Andrade & Du, 2005; Goodrich, 1997). Popham (2007) dereceli puanlama anahtarlarının; değerlendirme ölçütleri, ölçüt tanımları ve bir puanlama stratejisi olmak üzere üç bölümden oluştuğunu ifade etmiştir. Değerlendirme ölçütleri; öğrenci performansının niteliğine karar verirken değerlendiricinin dikkate aldığı kriterlerdir. Bu kriterler öğrencilerin başarılı olması için karşılanması gereken durumlardır (Wiggins, 1991). Ölçüt tanımları; öğrencilerin göstermesi gereken performans düzeyleri ile ilgili detaylı tanımlardır (Popham, 2007). Puanlama stratejileri ise; puanlamanın sürece mi yoksa sonuca mı dönük yapılacağını belirlemeyi içerir (Moskal, 2000). Dereceli puanlama anahtarı holistik ve analitik olmak üzere iki türde oluşturulmaktadır. Öğrencilerin gösterdiği performans veya ürün bir bütün olarak ele alınıp değerlendirildiğinde holistik dereceli puanlama anahtarı; öğrencilerin gösterdiği performans veya ürün ile ilgili bileşenler ayrı ayrı değerlendirildiğinde ise analitik dereceli puanlama anahtarı kullanılır (Atılğan, Kan, & Doğan, 2007).

Dereceli puanlama anahtarları üst düzey becerilerin değerlendirilmesi amacına hizmet ettiği için son yıllarda oldukça popüler araçlar haline gelmiştir (Hafner & Hafner, 2003). Bu araçlarla öğrencilerin üst düzey düşünme becerilerinin geliştirilmesine olanak sağlanır (Halonen ve diğ., 2003); böylece öğrenmeleri desteklenir ve artırılır (Andrade, 2000; Andrade & Boulay, 2003). Dereceli puanlama anahtarı kullanmanın en önemli amacı öğrencilere öğrenme ile ilgili beklentilerin açıkça ifade edilmesidir (Luft, 1997; 1999). Yani öğretmenin beklentilerini öğrenciye iletmeye yardımcı olur. Bu nedenlerden dolayı dereceli puanlama anahtarının, sadece değerlendirme aracı olmadığı aynı zamanda öğrenme aracı da olduğu söylenebilir. Dereceli puanlama anahtarı kullanımının değerlendirme süreci ile ilgili amaçlarına baktığımızda öğrencilerin kendi öğrenmelerini değerlendirme konusunda sorumluluk duymalarına (Phillip, 2002) ve çalışmalarını ile ilgili kendilerini değerlendirme imkanı bularak değerlendirme sürecine aktif katılmalarına yardımcı olduğu bildirilmektedir (Whittaker, Spencer, & Duhaney, 2001). Ayrıca öğrenciler dereceli puanlama anahtarı sayesinde kendi ihtiyaçlarını belirleme ve performanslarının zayıf ve güçlü olduğu noktalarda kendilerini değerlendirme imkanı bulabilmektedirler (Andrade & Du, 2005). Dereceli puanlama anahtarları çeşitli şekillerde oluşturulabildiği için pek çok ders alanı için kolayca kullanılabilir (Moskal, 2000). Özellikle proje ve sözlü sunum gibi uygulama temelli alanlara not vermede, öğretmenlerin yaşadıkları memnuniyetsizlikten dolayı kendisine kullanım alanı bulmuştur (Reddy, 2007).

Sunum yapma, etkili öğretim için gerekli önemli unsurlardan biri olması sebebiyle öğretmen yetiştiren programlarda mutlaka geliştirilmesi gereken mesleki beceriler arasındadır. Etkili bir sözlü sunum performansı, içeriği planlama ve organize etme, etkili sözlü ve görsel etkileşim, anlamlı beden dili ve zamanı verimli kullanma gibi pek çok alt beceriyi gerektiren çok yönlü bir performanstır. Öğretmen adaylarının sözlü sunum becerilerinin farkına varmalarının sağlanması ve geliştirmelerine yönelik fırsatlar yaratılması, mesleki yaşamlarında etkili öğretmenler olabilmeleri için önemli katkılar sağlayabilir. Bu çalışmada öğretmen adaylarının sözlü sunum performanslarının, dereceli puanlama anahtarı ile akranları, kendileri ve öğretmenleri tarafından değerlendirilmesi sonrasında öz, akran ve öğretmen değerlendirmelerinin karşılaştırılması amaçlanmaktadır. Öz, akran ve öğretmen değerlendirmeleri arasındaki olası farklılıkların incelenmesi sonucunda, öğretmen adaylarının performans değerlendirmeye ilişkin yeterliklerinin ortaya konulması hedeflenmektedir. Çalışmanın, bu doğrultuda yapılabilecek iyileştirmelere ilişkin bir çerçeve çizilmesiyle bu becerilerin geliştirilmesine katkı sağlaması beklenmektedir.

## Yöntem

Öğretmen adaylarının performanslarına ilişkin öz, akran ve öğretmen değerlendirmelerinin karşılaştırılması amacıyla gerçekleştirilen bu çalışmada var olan bir durum ortaya konulmaya çalışıldığından betimsel araştırma yöntemi izlenmiştir. Çalışma kapsamında öğretmen adayları dereceli puanlama anahtarı ile kendi performansları ve sınıf arkadaşlarının performanslarını değerlendirirken, öğretmenleri tarafından da değerlendirilmişlerdir. Değerlendirmeler tamamlandıktan sonra her bir öğretmen adayına ait öz, akran ve öğretmen değerlendirme puanları karşılaştırılmıştır. Ulaşılan sonuçlar doğrultusunda öğretmen adaylarının öz, akran ve performans değerlendirme konularındaki yeterliklerine ilişkin mevcut durum değerlendirilmiştir.

### Çalışma grubu

Araştırmanın çalışma grubunu, 2013-2014 eğitim öğretim döneminde bir devlet üniversitesinde öğrenim gören 18 öğretmen adayı (13 kız, 5 erkek) ve 1 alan eğitimi uzmanı oluşturmuştur. Çalışma, Eğitim Fakültelerinin 2. sınıfında öğretimi gerçekleştirilen 'Öğretim İlke ve Yöntemleri' dersi kapsamında yürütülmüştür. Bu dersin içeriğinin önemli bir bölümünü öğretim yöntem ve teknikleri ile bunların uygulama ile ilişkisi oluşturmaktadır. Bu bağlamda ders kapsamında öğretmen adaylarının, içeriğinde öğretim yöntem ve tekniklerinin uygulamasının yer aldığı sunumlar gerçekleştirmeleri sağlanmıştır. Sözlü sunumlar yaklaşık olarak 30 dakika sürmüştür ve her bir öğretmen adayı farklı bir konu içeriğini bir öğretim yöntem veya tekniğini uygulayarak sunmuştur. Sunumların öncesinde, öğretim yöntem ve tekniklerinin uygulanmasına ilişkin çeşitli örnekler araştırmacılar tarafından gösterilmiş ve gerekli açıklamalarda bulunulmuştur. Ayrıca çalışmaya katılan öğretmen adayları daha öncesinde etkili sunum teknikleri konusunda araştırmacılar tarafından bilgilendirilmişlerdir. Bunun yanı sıra çalışma grubunda yer alan öğretmen adaylarının tamamı öz değerlendirme ve akran değerlendirme konusunda deneyime sahiptirler.

### Verilerin toplanması

Çalışmanın verileri 8 haftalık sürede toplanmıştır. Sözlü sunumları sonrasında her bir öğretmen adayının performansı analitik türde geliştirilmiş dereceli puanlama anahtarı aracılığıyla değerlendirilmiştir. Öğretmen adaylarının sunum performansı konusunda kendisinden beklenenleri tam olarak bilmesi için dereceli puanlama anahtarı önceden öğrencilere sunulmuştur. Ayrıca dereceli puanlama anahtarının kullanımı ve puanlamasına yönelik detaylı açıklamalar yapılmıştır.

Araştırmada kullanılan analitik dereceli puanlama anahtarı geliştirilirken, Andrade (2001) tarafından önerilen basamaklar dikkate alınmıştır. Çalışmada izlenen bu basamaklar aşağıda sunulmuştur:

1. Performansı belirlemede kullanılacak ölçütlerin belirlenmesi
2. Kullanılacak dereceli puanlama anahtarına karar verilmesi
3. Performans düzeylerinin belirlenmesi ve düzey tanımlamalarının yapılması
4. Uzman görüşünün alınması

Dereceli puanlama anahtarı geliştirilirken ölçütlerin belirlenmesi aşamasında ilk olarak, etkili öğretimin gerçekleştirilebilmesi için sunum performanslarında bulunması gereken özelliklerle ilgili araştırmalar incelenerek dersin hedefleri ile örtüşen ölçütlere karar verilmiştir. Sonrasında çalışmaya katılan öğretmen

adaylarının da görüşleri dikkate alınarak toplam altı sözlü sunum ölçütü belirlenmiştir. Bu ölçütlerden iki tanesi içerik ile ilgili (içeriğin organizasyonu, içeriğin özgünlüğü); üç tanesi anlatım ile ilgili (ifade gücü, beden dili, hakimiyet); bir tanesi zaman kullanımı ile ilgilidir. Öğrencilerin sözlü sunum performanslarının bileşenleri, ayrı ayrı incelenmek istendiğinden analitik türde dereceli puanlama anahtarı kullanılmasına karar verilmiştir. Belirlenen ölçütler, puanlama matrisi içerisinde düzenlenmiş ve her bir ölçüt için puanlama düzeyinin keşiştiği gözeneğe o düzeyde öğrencinin göstermesi gereken performans iyiden kötüye (4'ten 0'a) doğru ayrıntılı şekilde tanımlanmıştır. Hazırlanan analitik dereceli puanlama anahtarına gerekli açıklayıcı yönergeler eklenerek, etkili ve objektif bir ölçme aracı haline getirilmesi sağlanmıştır. Sözlü sunum performanslarıyla ilgili değerlendirme 4-0 arasında değişen başarı düzeylerine göre puanlanmıştır. Geliştirilen analitik dereceli puanlama anahtarının (Şekil 1.) geçerliği konusunda uzman görüşüne başvurulmuş ve üzerinde gerekli düzenlemeler yapılarak uygulamaya hazır hale getirilmiştir.

ÖLÇÜTLER		4 (Mükemmel)	3 (Başarılı)	2 (Geliştirilmesi gerekli)	1 (Yetersiz)	0 (Başarısız)	Puan
İÇERİK	İçeriğin organizasyonu	Konuya dair içerik mükemmel bir biçimde sentezlenmiş.	Konuya dair içerik çoğunlukla doğru bir biçimde sentezlenmiş.	Konuya dair içerik kısmen doğru bir biçimde sentezlenmiş.	Konuya dair içerik çok az doğru bir biçimde sentezlenmiş.	Konuya dair içerik sentezlenmemiş.	
	İçeriğin özgünlüğü	Konunun içeriği tamamen özgündür.	Konunun içeriği çoğunlukla özgündür.	Konunun içeriği kısmen özgündür.	Konunun içeriğinin çok azı özgündür.	Konunun içeriği özgün değildir.	
ANLATIM	İfade gücü	Konu, dinleyicinin tamamen anlayabileceği, ilgisini çekebileceği şekilde sunuldu.	Konu, dinleyicinin çoğunlukla anlayabileceği, ilgisini çekebileceği şekilde sunuldu.	Konu, dinleyicinin kısmen anlayabileceği, ilgisini çekebileceği şekilde sunuldu.	Konu, dinleyicinin çok az anlayabileceği, ilgisini çekebileceği şekilde sunuldu.	Konu, dinleyicinin anlayabileceği i, ilgisini çekebileceği şekilde sunulmadı.	
	Beden dili	Beden dili tamamen doğru bir biçimde kullanıldı.	Beden dili çoğunlukla doğru bir biçimde kullanıldı.	Beden dili kısmen doğru bir biçimde kullanıldı.	Beden dili çok az doğru bir biçimde kullanıldı.	Beden dili doğru bir biçimde kullanılmadı.	
	Hakimiyet	Konuya tamamen hakimiyet sağlandı.	Konuya çoğunlukla hakimiyet sağlandı.	Konuya kısmen hakimiyet sağlandı.	Konuya çok az hakimiyet sağlandı.	Konuya hakimiyet sağlanamadı.	
ZAMAN KULLANIMI		Konuyu verilen süre içinde tamamladı.	Verilen süreye +/- 0-3 dakika uymadı.	Verilen süreye +/- 3-6 dakika uymadı.	Verilen süreye +/- 6-9 dakika uymadı.	Verilen sürenin oldukça altında veya üstünde sunumu tamamladı.	
<b>Toplam</b>							

Şekil 1. Sunum becerileri dereceli puanlama anahtarı

Öğretmen adayları sözlü sunum performanslarını, belirlenen altı ölçüte göre 0 ve 4 aralığında değişen puanlarla derecelendirmişlerdir. Bu durumda alınabilecek en düşük puan 0, en yüksek puan 24'tür. Her bir öğretmen adayının performansı 17 akranı tarafından değerlendirilmiştir. Dannefer ve arkadaşları (2005) mesleki yeterlilikleri ölçmede orta düzeyde bir güvenilirliğin sağlanması için en az altı akranı gerektirdiğini ifade etmektedirler. Araştırma deseni gereği öğretmen adaylarının sunum performansları ayrıca kendileri ve alan eğitimi uzmanı olan bir öğretmen tarafından, çalışma için geliştirilmiş analitik dereceli puanlama anahtarı kullanılarak değerlendirilmiştir. Öz, akran ve öğretmen değerlendirmeleri birbirinden bağımsız olarak gerçekleştirilmiştir.

### Verilerin analizi

Analitik dereceli puanlama anahtarlarından elde edilen verilerin analizi SPSS 20 programı kullanılarak gerçekleştirilmiştir. Öğretmen adaylarının sözlü sunum performanslarına ilişkin öz, akran ve öğretmen değerlendirmesinden elde ettikleri puanların, ortalama, standart sapma, minimum ve maksimum değerleri ile basıklık ve çarpıklık katsayılarını içeren betimsel istatistikleri belirlenmiştir. Basıklık ve çarpıklık katsayıları incelendiğinde ölçümlere ait verilerin normallikten sapma gösterdiği görülmüştür. Ayrıca verilerin, normallik koşulunun sağlanıp sağlanmadığının kontrolü için Shapiro-Wilk testi ile analizi yapılmıştır. Sonuçlar, öz değerlendirmeye ilişkin veriler dışında diğer ölçümlere ilişkin verilerin normal dağılım göstermediğini ortaya koymuştur (Öz değerlendirme ölçümü için Shapiro-Wilk testi sonuçları = 0.931,  $df = 18$ ,  $p = .206$ ; Akran değerlendirme ölçümü için Shapiro-Wilk testi sonuçları = 0.786,  $df = 18$ ,  $p = .001$ ; Öğretmen değerlendirme ölçümü için Shapiro-Wilk testi sonuçları = 0.767,  $df = 18$ ,  $p = .001$ ). Bu yüzden, verilerin çözümlenmesinde parametrik olmayan testler ile analizler gerçekleştirilmiştir. Öz, akran ve öğretmen değerlendirmelerinin karşılaştırılmasına ilişkin araştırma probleminin çözümlenmesi için Friedman testi gerçekleştirilerek, öz, akran ve öğretmen değerlendirmelerine ilişkin puanlar arasında istatistiksel olarak anlamlı farklılıklar bulunup bulunmadığı araştırılmıştır. Farkların hangi ölçümler arasında olduğunu ortaya koymak için Wilcoxon işaretli sıralar testi ile ikili karşılaştırmalar yapılmıştır. Analiz sonuçlarına ilişkin etki büyüklük değerleri hesaplanarak rapor edilmiştir.

### Bulgular

Öz, akran ve öğretmen değerlendirme puanlarına ilişkin ortalama, standart sapma, en düşük ve en yüksek değerler ile basıklık ve çarpıklık katsayılarını içeren betimsel istatistikler Tablo 1'de gösterilmiştir.

Tablo 1. Öz, Akran ve Öğretmen Değerlendirme Puanlarına Ait Betimsel İstatistikler

	Öz değerlendirme	Akran değerlendirme	Öğretmen değerlendirme
<b>Ortalama</b>	17,33	20,83	19,78
<b>Standart sapma</b>	3,32	2,04	4,17
<b>En düşük puan</b>	12	15	6
<b>En yüksek puan</b>	24	23	24
<b>Çarpıklık</b>	0,632	-1,185	-2,261
<b>Çarpıklığın standart hatası</b>	0,536	0,536	0,536
<b>Basıklık</b>	-0,295	3,410	6,876
<b>Basıklığın standart hatası</b>	1,038	1,038	1,038

Tablo 1'de görülen değerlendirmelere ait ortalama puanlar incelendiğinde, öğretmen adaylarının sözlü sunum performanslarına ilişkin akran değerlendirme puanları ortalamasının ( $\bar{x} = 20,83$ ), öz değerlendirme ( $\bar{x} = 17,33$ ) ve öğretmen değerlendirme ( $\bar{x} = 19,78$ ) puanları ortalamalarından daha yüksek olduğu görülmektedir. Ortalama puanlar arasında görülen bu farklılıkların istatistiksel açıdan anlamlı olup

olmadığını belirlemek üzere Friedman testi gerçekleştirilmiştir. Öğretmen adaylarının öz, akran ve öğretmen değerlendirmesine ilişkin karşılaştırma sonuçları Tablo 2'de sunulmuştur.

Tablo 2. Öğretmen Adaylarının Öz, Akran ve Öğretmen Değerlendirmelerinin Karşılaştırılmasına İlişkin Friedman Testi Sonuçları

	N	$\chi^2$	sd	p
Öz değerlendirme -				
Akran değerlendirme -	18	14,576	2	.001
Öğretmen değerlendirme				

Tablo 2'de görüldüğü üzere çalışma grubunda yer alan 18 öğretmen adayının öz, akran ve öğretmen değerlendirmesi arasındaki farkın karşılaştırılması amacıyla gerçekleştirilen Friedman testinin sonucunda, öz, akran ve öğretmen değerlendirmesi arasındaki farkın istatistiksel olarak anlamlı olduğu sonucuna ulaşılmıştır ( $\chi^2_{(2,N=18)} = 14,576, p < 0.05$ ). Değişkenlere ilişkin sıra ortalamaları incelendiğinde akran değerlendirme (2,53), öz değerlendirme (1,33) ve öğretmen değerlendirmesinden (2,14) yüksek bulunmuştur. Bu farkın kaynağını belirleyebilmek için Wilcoxon işaretli sıralar testi yapılmış ve sonuçları Tablo 3'de sunulmuştur.

Tablo 3. Öğretmen Adaylarının Öz, Akran ve Öğretmen Değerlendirmelerinin Karşılaştırılmasına İlişkin Wilcoxon İşaretli Sıralar Testi Sonuçları

		N	Sıra Ortalaması	Sıra Toplamı	z	p
<b>Akran değerlendirmesi- Öz değerlendirme</b>	Negatif sıralar	2	5,25	10,50	-2,983	0,003
	Pozitif sıralar	14	8,96	125,50		
	Eşit sıralar	2				
	Toplam	18				
<b>Öz değerlendirme- Öğretmen değerlendirmesi</b>	Negatif sıralar	14	8,14	114,00	-2,396	0,017
	Pozitif sıralar	2	11,00	22,00		
	Eşit sıralar	2				
	Toplam	18				
<b>Akran değerlendirmesi- Öğretmen değerlendirmesi</b>	Negatif sıralar	5	8,50	42,50	-1,640	0,101
	Pozitif sıralar	12	9,21	110,50		
	Eşit sıralar	1				
	Toplam	18				

Tablo 3'de görüldüğü gibi Wilcoxon işaretli sıralar testi sonucuna göre öz değerlendirme ve akran değerlendirmesi arasındaki fark istatistiksel olarak anlamlıdır ( $z = -2,983, p \leq 0,05$ ). Fark puanların pozitif sıralar lehine olması, akran değerlendirmesinin öz değerlendirmeden anlamlı biçimde yüksek olduğunu göstermektedir. Anlamlı bulunan bu farkın etki büyüklük değeri hesaplanarak 0,50 bulunmuştur, bu değer orta düzeyde bir etki büyüklüğünü ifade etmektedir (Field, 2005). Tablo 3'de görülen bir diğer istatistiksel olarak anlamlı sonuç, öz değerlendirme ve öğretmen değerlendirmesi puanlarının karşılaştırılmasına ilişkindir. Buna göre öğretmen adaylarının öz değerlendirme ile öğretmen değerlendirmesi puanları ortalamaları arasındaki fark anlamlı bulunmuştur ( $z = -2,396, p \leq 0,05$ ). Sıra ortalamaları ve sıra toplamları incelendiğinde fark puanların negatif sıralar lehine olduğu görülmektedir. Bu durum öğretmen değerlendirmesinin öz değerlendirmeden anlamlı biçimde yüksek olduğunu göstermektedir. Bu farkın etki büyüklüğü 0,40 olarak hesaplanmıştır ve bu değer orta düzeyde bir etki büyüklüğüne işaret etmektedir (Field, 2005). Tablo 3'de görülen bir diğer sonuca göre, Wilcoxon işaretli sıralar testi, akran

değerlendirmesi ve öğretmen değerlendirme puan ortalamaları arasındaki farkın istatistiksel olarak anlamlı olmadığını göstermiştir ( $z = -1,640, p > 0,05$ ).

## Tartışma ve Sonuç

Öğretmen eğitiminde kazandırılması kaçınılmaz olan öz düzenleme, eleştirel düşünme ve karar verme gibi becerilerin geliştirilmesi için öz değerlendirme ve akran değerlendirme uygulamaları önemli fırsatlar sunmaktadır. Sluijsmans, Brand-Gruwel ve van Merriënboer (2002), özellikle öğretmen eğitimi kapsamında, öğrencilerin akranlarının performanslarını eleştirel olarak değerlendirebilmeleri konusunda eğitilmelerinin önemli olduğunu vurgulamaktadırlar. Bu çalışmada öğretmen adayları ile öz, akran ve öğretmen değerlendirme uygulamaları gerçekleştirilmiş ve değerlendirme sonuçları karşılaştırılmıştır.

Çalışmada elde edilen sonuçlar, öğretmen adaylarının öz değerlendirmelerinin, akran ve öğretmen değerlendirmesinden farklılık gösterdiğini ortaya koymuştur. Öğretmen adaylarının performanslarına ilişkin olarak yaptıkları öz değerlendirmelerin, akranları ve öğretmenleri tarafından yapılan değerlendirmelerden anlamlı olarak düşük olduğu belirlenmiştir. Bu sonuç öğretmen adaylarının kendi performanslarına ilişkin olarak, akranlarından ve öğretmenlerinden çok daha eleştirel olduklarının bir göstergesi olarak ortaya çıkmaktadır. Akran değerlendirmesinin öz değerlendirmeden yüksek oluşunun nedenlerinden biri, öğretmen adaylarının arkadaşlarını değerlendirirken kendilerini rahat hissetmemeleri olabilir. Öğretmen adaylarının, arkadaşlarını düşük puan vererek cezalandırmaktan çekindikleri için yüksek puanlar verdikleri düşünülmektedir. Eğitimlerinin erken kademelerinden itibaren bu tür uygulamalar konusunda deneyim edinmemiş öğrencilerin, kendilerini akran ve öğretmen değerlendirmelerine göre farklı düzeyde değerlendirmeleri beklenen bir sonuç olarak ortaya çıkmıştır. Akran değerlendirme ile öğretmen değerlendirmesinden elde edilen puanlar karşılaştırıldığında ise akran ve öğretmen değerlendirme arasında anlamlı bir farkın görülmediği sonucuna ulaşılmıştır. Bu sonucun, dereceli puanlama anahtarında ölçütlerin ve performans tanımlamalarının açık ve net ifade edilmiş olmasının objektif ve tutarlı değerlendirmelere olanak sağlaması ile ilişki olduğu düşünülebilir.

Yüksek öğretimde öz, akran ve öğretmen değerlendirmesinin kullanılmasına ilişkin olarak yürütülen araştırmaların derlendiği bir çalışmada öz, akran ve öğretmen değerlendirmelerinin karşılaştırma sonuçlarının önemli ölçüde farklılaştığı ifade edilmekle beraber öz, akran ve öğretmen değerlendirmesinin birlikte kullanılmasının öğrencilerin öğrenmelerinin niteliğini farklı açılardan iyileştirdiği vurgulanmaktadır (Dochy ve diğ., 1999). Bu çalışmada elde edilen sonuçlar da bazı araştırma sonuçları ile tutarlılık gösterirken bazılarında farklılaşmaktadır (Dochy ve diğ., 1999; Sluijsmans ve diğ., 1999; Somervell, 1993). Bu duruma yol açan neden araştırmaların çalışma gruplarında yer alan katılımcı sayısının farklılığı olabileceği gibi, katılımcıların geçmişte öz değerlendirme ve akran değerlendirme konusundaki deneyimleri ile de ilgili olabilir. Ayrıca çalışmada, araştırma sürecinde elde edilen sonuçların genellenebilir olmasından ziyade öğretmen adaylarının performans değerlendirmeye ilişkin yeterliklerinin geliştirilmesine yönelik bir süreç izlenmiştir. Sonuç olarak da çalışma kapsamında gerçekleştirilen uygulamalar ile öğretmen adaylarının performanslarının daha nitelikli olması ayrıca öz ve akran değerlendirme konusunda deneyim sahibi olmaları sağlanmıştır.

Genel bir sonuç olarak, öz değerlendirme ve akran değerlendirme uygulamalarının çeşitli sınırlılıklarının yanı sıra, öğrencileri motive etme, öğrenmelerinin farkında olmalarını sağlama ayrıca öğrenme süreç ve ürünlerinin niteliğini iyileştirme gibi olumlu özelliklere sahip olduğu görülmektedir. Öğretmen adaylarının kendilerini eleştirel göze değerlendirebilmeleri ve akranlarının performansını doğru ve nesnel olarak değerlendirebilmeleri konusunda geliştirilmesi ileride meslek yaşantılarında objektif bir tutum sergilemelerine katkı sağlayacaktır. Bu nedenlerle öz ve akran değerlendirme uygulamalarına ölçme ve



değerlendirme sürecinde daha sık yer verilmesinin ve sonuçlarının araştırılmasının öğretmen eğitimine önemli katkılar sağlayacağı düşünülmektedir. Ayrıca öğretim ortamlarında öz ve akran değerlendirme uygulamalarına eğitimin erken kademelerinden itibaren yer verilmesi ile, öğrencilerin bu konularda deneyim sahibi olmaları ile beraber daha nesnel ve tutarlı değerlendirmeler yapabilme becerilerinin geliştirilmesi yönünde olanak sağlanabileceği düşünülmektedir. Bununla beraber dereceli puanlama anahtarının sözlü sunum gibi performansa dayalı değerlendirmelerde kullanılması sonucunda yapılan değerlendirmeler öz, akran ve öğretmen değerlendirme becerilerinin gelişimine katkı sağlayacaktır.

Araştırmadan elde edilen sonuçlara dayanarak öğretim sürecine ve sonraki araştırmalara yönelik olarak şu öneriler ifade edilebilir: Bir performansa yönelik öğretmen adaylarına neler yapması gerektiğiyle ilgili dikkat etmeleri gereken noktaların dereceli puanlama anahtarları kullanılarak belirtilmesi öğretmen adaylarının algılarının ve becerilerinin daha iyi hale gelmesine katkı sağlayabilir. Dereceli puanlama anahtarının değerlendirme sürecindeki öneminin vurgulanması ayrıca öğretmen adaylarının dereceli puanlama anahtarı ile ilgili bilgilendirilerek, geliştirilmesinin ve kullanımının yaygınlaştırılması okul eğitimine katkı sağlayabilir. Sözlü sunum dışında farklı beceri alanlarına ilişkin yeni dereceli puanlama anahtarlarının geliştirilerek, geniş ve çeşitli gruplar üzerinde uygulanması ile üst düzey becerilerin iyileştirilmesi ve tamamlayıcı değerlendirme araçlarının etkileri açısından tartışılması önerilebilir.

### Kaynakça

- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57 (5), 13-18.
- Andrade, H. G. (2001). The effects of instructional rubrics on learning to write. *Current Issues in Education [On-line]*, 4 (4). <http://cie.asu.edu/volume4/number4/> adresinden 12 Ekim 2010 tarihinde indirilmiştir.
- Andrade, H. G. (2005). Teaching with rubrics: the good, the bad, and the ugly. *College Teaching*, 53 (1), 27-31.
- Andrade, H. G. & Boulay, B. A. (2003). Role of rubric-referenced self-assessment in learning to write. *The Journal of Educational Research*, 97 (1), 21-30.
- Andrade, H. G. & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research and Evaluation*, 10 (3), 1-11.
- Atılğan, H., Kan, A., & Doğan, N., (2007). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı Yayıncılık.
- Berberoğlu, G. (2006). *Sınıf içi ölçme ve değerlendirme teknikleri*. İstanbul: Morpa Yayıncılık.
- Bostock, S. (2001). *Student peer assessment*. In Higher Education Academy [http://www.heacademy.ac.uk/resources/detail/resource\\_database/id422\\_student\\_peer\\_assessment](http://www.heacademy.ac.uk/resources/detail/resource_database/id422_student_peer_assessment) adresinden 10 Ekim 2013 tarihinde indirilmiştir.
- Boud, D. & Falchikov, N. (1989). Quantitative studies of self-assessment in higher education: a critical analysis of findings. *Higher Education*, 18, 529-549.
- Dannefer, E., Henson, L., Bierer, S., Grady-Weliky, T., Meldrum, S., Nofziger, A., et al. (2005). Peer assessment of professional competence. *Medical Education* 39 (7), 713-722.

- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: a review. *Studies in Higher Education*, 24 (3), 331–350.
- Falchikov, N. (1995). Peer feedback marking: developing peer assessment. *Innovations in Education and Training International*, 32, 175-187.
- Field, A. (2005). *Discovering statistics using SPSS* (2nd ed.). London: Sage Publications.
- Goodrich, H. G. (1997). Understanding rubrics. *Educational Leadership*, 54 (4), 14-17.
- Hafner, J. C., & Hafner, P. M. (2003). Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education*, 25 (12), 1509-1528.
- Halonen, J. S., Bosack, T., Clay, S., McCarthy, M., Dunn, D. S., Hill IV, G., W., et al. (2003). A rubric for learning, teaching, and assessing scientific inquiry in psychology. *Teaching of Psychology*, 30 (3),196-208.
- Jensen, K. (1995). Effective rubric design. *The Science Teacher*, 62 (5), 34-37.
- Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2008). *Öğrenci başarısının belirlenmesi: performansa ve portfolyoya dayalı durum belirleme*. Ankara: Pegem Akademi.
- Lindblom-Ylänne, S., Pihlajamäki, H., & Kotkas, T. (2006). Self-, peer- and teacher-assessment of student essays. *Active Learning in Higher Education*, 7 (1), 51–62.
- Luft, J. A. (1997). Design your own rubric. *Science Scope*, 20 (5), 25-27.
- Luft, J. A. (1999). Rubrics: design and use in science teacher education. *Journal of Science Teacher Education*, 10 (2), 107-121.
- MEB (2007). *Ortaöğretim 9.sınıf biyoloji dersi öğretim programı*. Ankara.
- Moskal, B. M. (2000). Scoring rubrics: what, when and how? *Practical Assesment, Research & Evaluation*, 7 (3). <http://PAREonline.net/getvn.asp?v=7&n=3> adresinden 24 Ekim 2013 tarihinde indirilmiştir.
- Phillip, C. (2002). Clear expectations: rubrics and scoring guides. *Knowledge Quest*, 31(2), 26-27.
- Popham, J. W. (2007). *Classroom assessment: what teachers need to know* (5th ed.). USA: Pearson Education.
- Reddy, M. (2007). Effect of rubrics on enhancement of student learning. *Educate*, 7 (1), 3-17.
- Saito, H. & Fujita, T. (2004). Characteristics and user acceptance of peer rating in EFL writing classrooms. *Language Teaching Research*, 8, 31–54.
- Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999). Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research*, 1, 293–319.

Sluijsmans, D., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: effects on performance and perceptions. *Assessment & Evaluation in Higher Education*, 27 (5), 443–454.

Somervell, H. (1993). Issues in assessment, enterprise and higher education: the case for self, peer and collaborative assessment. *Assessment & Evaluation in Higher Education*, 18, 221–233.

Topping, K. J. (2009). Peer assessment. *Theory into Practice*, 48 (1), 20–27.

Whittaker, R. C., Spencer, S. J., & Duhaney, D. (2001). Creating instructional rubrics for inclusive classroom. *Teaching Exceptional Children*, 34 (2), 8-13.

Wiggins, G. (1991). Standart, not standardization: evoking quality student work. *Educational Leadership*, 48 (5), 18-25.

### **Extended Abstract**

As a result of the reflections of new teaching approaches onto learning environments, enhancing students' scientific process skills and training of students as creative and critical, inquisitive and questioning individuals have gained greater importance and students are encouraged to be ready for real life situations. All these new tendencies affect in-class evaluation and assessment as well as all the components of education. While the existing evaluation and assessment tools are effective in making decisions about whether students have internalized the given information, performance-based evaluation instruments are needed to determine the extent to which they can apply their knowledge in real situations. In recent years, rubrics have gained greater priority among performance evaluation tools to evaluate student performance. Moreover, self- and peer-assessment applications have aroused great deal of interest in educational research.

As self- and peer-assessment have advantages complementary to each other, they are frequently employed together in instructional settings. Researchers stress that self- and peer-assessment can be effective in motivating students, encouraging them to actively participate, ensuring effective learning to occur, supporting students to take their responsibility and control for their own learning and helping them to construct their learning better. Use of specific criteria, transparent evaluation processes, clear and understandable instructions, use of scoring matrices are claimed to make contributions to self- and peer-assessment applications; moreover, rubrics are recommended to be used for self- and peer-assessment.

As rubrics consider the use of higher level cognitive skills, they have become quite popular in recent years. The main cause of using a rubric is its allowing the clear expression of expectations related to learning to students. In addition, through a rubric, students are provided with opportunities to determine their own needs and evaluate the weak and strong aspects of their performance.

As presentation skill is one of the important elements needed for effective learning, it is among the professional skills to be developed in teacher training institutions. Making pre-service teachers aware of their oral presentation skills and enabling them to enhance these skills are of great importance for them to be effective teachers throughout their professional career. The present study aims to compare self-, peer- and teacher-assessment of the pre-service teachers' oral presentation skills through a rubric.

In the current study combining research and application in order to evaluate and improve the pre-service teachers' oral presentation skills, action research method was used as it makes the transfer of research

findings into practice easier. The study group of the current research consists of 18 pre-service teachers and 1 subject area education expert. The oral presentations made within the framework of the current study lasted about 30 minutes and each pre-service teacher made his/her presentation about a different topic by using an instructional method or technique. After the oral presentations, performance of each pre-service teacher was evaluated by means of the rubric developed in the analytic format. While developing the analytic rubric, the criteria were structured on the basis of the review of the literature focusing on the characteristics to be involved in a presentation performance for effective instruction to occur and on the basis of the pre-service teachers' opinions. In this way, totally six oral presentation performance criteria were determined. Two of the criteria are related to content (organization of the content, authenticity of the content); three of them are related to expression (power of expression, body language and mastery of the issue); one of them is related to the use of time. The pre-service teachers' oral presentation performances were evaluated by assigning scores ranging from 0 to 4 according to the above-mentioned six criteria. Each pre-service teacher's performance was evaluated by his/her 17 peers. The pre-service teachers' oral presentation performances were also evaluated by themselves and an expert by using the analytic rubric developed for the study. In order to compare the self-, peer- and teacher-assessment, Friedman and Wilcoxon signed rank tests was run.

The analysis of the data revealed that there is a statistically significant difference between means of self- and peer-assessment scores. When the difference was analyzed, it was found that peer-assessment is significantly higher than self-assessment. Another finding of the data analysis showed that the difference between the mean scores of self- and teacher-assessment is statistically significant. When the rank means and rank sums were examined, it was found that the teacher-assessment is significantly higher than self-assessment. Moreover, Wilcoxon signed rank test showed that the difference between mean scores of peer- and teacher-assessment is not statistically significant.

Pre-service teachers' self-assessments of their oral presentation performances are found to be significantly lower than peer- and teacher-assessments of their performances. This shows that the pre-service teachers are more critical of their own performances than their peers and teachers. One of the important reasons for peer assessments' being higher than self-assessments is that they may have felt stressed while evaluating their peers. The pre-service teachers may have given higher scores to their peers as they may have desired to avoid punishing their peers by giving low scores. When the scores obtained from peer- and teacher-assessments were compared, no significant difference was observed. This may be because of the fact that clear and understandable expression of the criteria and performance descriptions included in the rubric allowed objective and consistent evaluations.

As a conclusion, within the context of the current study, the pre-service teachers' oral presentations were made more qualified and they were provided with experiences about self- and peer-assessment. When pre-service teachers can critically evaluate themselves and the performances of their peers accurately and objectively, they will be better prepared to demonstrate objective attitudes in their future professional career. Thus, more frequent use of self- and peer-assessments and rubrics in assessment and evaluation process and investigation of their outcomes are believed to make important contributions to teacher education.