



## Dereceli Puanlama Anahtarı ve Puanlama Anahtarından Elde Edilen Puanların Uyum Düzeyleri \*

### Comparison of Answer Key and Scoring Rubric For the Evaluation of The Student Performances

Burcu PARLAK\*\*, Nuri DOĞAN\*\*\*

**ÖZ:** Bu çalışmada, mesleki bilgi gerektiren uygulamalı bir derste, puanlama anahtarı ve dereceli puanlama anahtarı (DPA-Rubrik) kullanılarak elde edilen puanların uyum düzeylerinin araştırılması amaçlanmıştır. Araştırma bir meslek lisesinin elektrik-elektronik bölümündeki 70 öğrenci ve 6 öğretmenin katılımı ile yapılmıştır. Elde edilen bulgulara göre, ölçme araçlarından elde edilen puanlar arasında yüksek bir uyum olduğu görülmüştür. Öğrencilerin 1.dönem notlarıyla, her iki ölçme aracı ile verilen puanlar arasındaki korelasyona bakılmış ve DPA ile verilen puanların korelasyonunun daha yüksek çıktığı görülmüştür. Ancak, iki korelasyon katsayısı farkı 0.05 düzeyinde manidar çıkmamıştır. Regresyon analizi sonucunda, DPA ile verilen puanların ders notlarını yordama düzeyi daha yüksek çıkmıştır.

**Anahtar sözcükler:** Puanlama Anahtarı, Analitik Dereceli Puanlama Anahtarı, Performans Değerlendirme.

**ABSTRACT:** In this study, it is aimed to analyze the adjustment levels of the points obtained by using scoring rubric and answer key for an applied course that requires technical knowledge. The researched was conducted in the electric & electronics major of a vocational high school on 70 students and 6 teachers. Based on the findings obtained it is seen that there is a high correlation between the points obtained by evaluation instruments. The correlation between the points of the students in the first semester and the scores awarded through both assessment methods has indicated that the correlation of the scores given through scoring rubric is higher. The correlation between the points for the first semester and the points obtained by using both scoring instruments were examined and it was seen that the correlation of the points obtained by scoring rubric is higher. However, the coefficient difference between two correlations was 0.05 and it is not significant. As a result of regression analyses, the regression level of points obtained by scoring rubric was higher than that of course points.

**Keywords:** Answer Key, Analytical Scoring Rubric, Performance Assessment

## 1. GİRİŞ

Öğrenci başarısını ölçmek için sıklıkla kullanılan yazılı, çoktan seçmeli, kısa yanıtlı gibi yöntemler daha çok öğrencilerin sınav becerilerine, sınav sonuçlarına ve bu sonuçlardan elde edilen sıralamalara odaklanmakta, öğrencinin düşünme becerilerini göz ardı etmektedir. Öğrenci performansının farklı durumlarda ve zamanla değiştiği göz önüne alındığında kısa bir zaman diliminde yapılan ve öğrenci başarısını grup içindeki yerine bakarak değerlendiren bu yöntemlerin özellikle süreci değerlendirmede yetersiz kaldığı söylenebilir. Öğretim programlarında ifade edilen öğrenci merkezli öğretim anlayışı, öğretmeni bir otorite merkezi olmaktan çıkarıp yol gösterici olmaya yönlendirmektedir. Öğrencinin öğrenme ortamına aktif katılımı ile bilgiye ulaşabilen ve bu bilgiyi kullanabilen bireylerin yetiştirilmesi esastır. Öğretim programlarında amaçlanan, öğrencileri öğrenme süreci içerisinde izleyerek ilgi ve yeteneklerinin ortaya çıkarılması, ürünün yanında sürecinde değerlendirilmesinin sağlanmasıdır. Bu farklı öğretim yöntemleri beraberinde yeni değerlendirme yöntemlerini de getirmiştir.

Bu yöntemlerden biri de performans değerlendirmedir. Performans değerlendirme, öğrencilerin gerçek hayatla ilgili sorunlara bilgi ve becerilerini kullanarak cevap verebilmelerini sağlayan, geçerliliği ve güvenilirliği yüksek ölçme araçlarıyla yapılan değerlendirme biçimi

\* Bu çalışma, ilk yazarın yüksek lisans tez çalışmasının bir kısmıdır.

\*\* Ölçme ve Değerlendirme Uzmanı, MEB, burcuprlk@yahoo.com

\*\*\* Doç.Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, nurid@hacettepe.edu.tr

olarak adlandırılabilir (Büyüköztürk 2007). Performans değerlendirmede, öğrenciden tek doğru cevabı bulması istenen sınav türlerinin yerine, cevabı kendisinin oluşturması veya bilgisini bir ürün ortaya koyarak göstermesi istenir (Stiggins 1994). McMillan (1997), objektif testler kullanarak öğretimin sonunda yapılan ölçmeler yerine, performansa dayalı ölçme yöntemleri kullanarak öğretim sırasında yapılan ölçmelerin önem kazandığını ifade ederken, Biemer, (1993) performans değerlendirme ile bilgilerin ne derece kazanıldığını ölçen bir yapıdan, bilgilerin yeni durumlarda ya da gerçek yaşam durumlarında nasıl kullanılabildiğini ölçen bir yapıya geçiş olduğunu belirtmektedir.

Bir işin yapılmasını gerektiren becerilerin bilgi testleri ile ölçülmesi zor olabilir. Eğitimde performans çoğu kez dolaylı olarak ölçülmektedir. Ancak, bu durum ölçme sonuçlarının geçerliliği açısından sakıncalıdır. Performansı belirleyen bilginin ölçülmesi, o işin nasıl yapılacağını bir kanıtı değildir. Çünkü bir işin nasıl yapılacağını bilmek ile işi belirlenen ölçütlere uygun olarak yapmak arasında her zaman yüksek bir ilişki yoktur. Bu nedenle, performansın ölçülmesinde, performansa uygun değerlendirme araçlarının kullanılması gerekmektedir (Tekin 1991).

Dereceli puanlama anahtarı (DPA) performans değerlendirmesinde kullanılan en yaygın ölçme araçlarından biridir. DPA ölçütlerden ve performans düzeylerinden oluşan yapısı ile puanlama sırasında oluşabilecek yanlışlıkları en aza indirirken, öğrencilere performanslarıyla ilgili daha gerçekçi ve ayrıntılı geri bildirim verme imkanı sağlar. İyi tanımlanmış bir DPA puanlamadan kaynaklanabilecek güvenilirlik problemlerini ortadan kaldırır.

DPA'yı Goodrich (2001), öğrenci durumunun belirlenmesinde ve izlenmesinde kullanılan puanlama kılavuzu olarak tanımlamıştır. Popham (1997), her bir çalışma için ölçütleri listeleyen ve çalışmada nelerin yapılacağını gösteren bir puanlama aracı olarak ifade ederken, Callison (2000), performansın kabul edilemez en düşük düzeyi ile gözlenebilir en yüksek düzeyi arasında belirlenen ölçütler dizisi olarak tanımlamıştır. Moskal'a (2000) göre, DPA öğrencilerin çalışmalarını veya ürünlerini analiz etmek için öğretmen rehberliğinde geliştirilmiş, tanımlanmış bir puanlama tasarımıdır ve konuların, etkinliklerin geniş bir aralıkta değerlendirilmesinde kullanılabilir. DPA; değerlendirme ölçütleri, ölçüt tanımlamaları ve bir puanlama stratejisi olmak üzere üç bölümden oluşur (Popham 1997; aktaran, Aslanoğlu ve Kutlu 2004).

*1. Değerlendirme ölçütleri:* Kabul edilebilir yanıtları, kabul edilemez yanıtlardan ayırmak için kullanılır. Örneğin, öğrencilerin açık uçlu sorulara verdikleri yanıtlar puanlanırken değerlendirilebilir ölçütler kullanılır.

*2. Ölçüt tanımlamaları:* Öğrencilerin yanıtlarındaki niteliksel farklılıkları tanımlama biçimini ifade eder. Örneğin bir kompozisyonda organizasyon değerlendirilecekse bu ölçütlerden en yüksek puanı alan öğrencinin kompozisyonu organizasyon açısından hiç hata içermemelidir.

*3. Puanlama stratejisi:* Puanlamada, bütünsel (holistic) ya da analitik (analytical) DPA kullanılabilir. Hangi tür DPA kullanılacağı değerlendirmenin amacına bağlıdır (Popham 1997). Bazı durumlarda performansın farklı düzeylerinin ortaya çıkarılması için belirlenmiş ölçütler arasında bir ayrışma bulunmamaktadır. Böyle durumlarda bütünsel puanlama anahtarı kullanılmalıdır (Brookhart 1999). Analitik puanlama anahtarı ise, ölçülen bir yetenek boyutunun öğelere ayrıştırılabildiğinde ve daha ayrıntılı puanlama yapmak istendiğinde kullanılmaktadır (Haladyna 1997).

Güvenirliği yüksek bir DPA geliştirmek için; ölçütler açık ve anlaşılır olmalıdır. Ölçütler birbirini kapsamamalı, her birinin ölçmesi gereken kriter farklı olmalıdır. Performans düzeyleri, öğrenciler arasındaki farklılıkları ortaya çıkaracak sayıda olmalıdır. Güvenilir bir puanlama için 3-5 arası bir performans düzeyi oluşturmak uygun görülmektedir. Haladyna (1997), düzeyin 7'ye kadar çıkarılmasını önermektedir.

Performans ölçmede kullanılabilir ve geleneksel olan bir diğer araç ise puanlama anahtarlarıdır (cetvelleridir). Bir soruya verilmesi beklenen yanıtlardan ölçütler listesi hazırlanır ve listedeki yanıtların her birine kaç puan verileceği belirlenerek puanlama cetveli oluşturulur. Puanlama cetveli öğretmenlere kılavuzluk edebilir ancak çeşitli nedenlerle (yanıtlayıcıyı tanıyarak olmak, dikkat eksikliği, ölçütlerin iyi tanımlanmamış olması vs.) puanlama yanlışlıkları söz konusu olabilir (Turgut 1997).

### 1.1. Araştırmanın Amacı

Müzik, resim, el sanatları, laboratuvar uygulamaları gibi performans ile ilgili hedefleri bulunan derslerde, kağıt kalem testleriyle ölçme değerlendirme yapmak oldukça zordur ve geçerlik sorunu taşımaktadır. Çünkü performans değerlendirmede, öğrenciden bir şeyin nasıl yapılacağını söylemesinden ziyade yapması istenir. Bundan dolayı, bu tür hedeflerin ölçülmesinde performans testleri kullanılmalıdır (Tekin 1991). Ancak performansa dayalı değerlendirmelerde, verilen puanların güvenilirliği konusunda yeterli özen gösterilmediği gözlenmektedir. Bu araştırma ile amaçlanan, uygulama gerektiren derslerde değerlendirme yaparken DPA'nın kullanılmasının öğrencilerin öğrenme süreçlerine getirdiği faydaları tespit etmek ve DPA ile elde edilen puanların daha geçerli ve güvenilir sonuçlar verip vermediğini belirlemektir. Bu amaçlara hizmet edeceği düşünülerek aşağıdaki sorulara yanıt aranmıştır.

1. "Puanlama anahtarı ve DPA ile puanlama yapan öğretmenlerin her bir grupta verdiği puanların uyum düzeyi nedir?"

2. "Puanlama anahtarı ve DPA ile elde edilen puanların ortalamaları arasında manidar bir fark var mıdır?"

3. "Puanlama anahtarı ve DPA'dan elde edilen puanlar ile öğrencilerin 1.dönem notları arasındaki korelasyonların farkı manidar mıdır?"

4. "Puanlama anahtarı ve DPA'dan elde edilen puanların, öğrencilerin 1.dönem notlarını yordama düzeyi nedir?"

Ayrıca bu çalışmanın analitik DPA geliştirme, uygulama ve sağladığı geri bildirimlerden yararlanma konusunda eğitimciler için kaynak olması, yapılan uygulamalarda ve öğrencilerin değerlendirilmesinde eksikliklerin giderilmesi konusunda katkı getirmesi beklenmektedir.

## 2. YÖNTEM

### 2.1. Araştırmanın Türü

Performans değerlendirmede DPA'nın kullanılması, elde edilen sonuçların incelenmesi ve puanlama anahtarından elde edilen sonuçlarla karşılaştırılması yapılmıştır. Buna benzer çalışmalar farklı derslerde de olsa daha önce yapılmış olup, bu çalışmadan elde edilen sonuçlar konuyla ilgili yapılacak sonraki araştırmalara katkı sağlayacağından yapılan bu çalışma temel araştırmalar sınıfına girmektedir.

### 2.2. Çalışma Grubu

Araştırmanın çalışma grubunu, Ankara'nın merkez ilçesinde bulunan Teknik ve Endüstri Meslek Lisesi'nin Elektrik Elektronik Bölümündeki 11. sınıf öğrencileri oluşturmaktadır. Araştırmaya üç sınıftan 70 öğrenci katılmıştır. Her sınıf uygulama çalışmalarında ikiye ayrıldığı için 6 grup oluşmuştur. Çalışma sürecinde her bir öğrencinin ayrı ayrı değerlendirildiği Analitik DPA, sınıf listelerinin bulunduğu ve öğrencilerin bu listeler üzerinden değerlendirilmesinin yapıldığı formlar (puanlama anahtarları) kullanılmıştır. Uygulama, puanlayıcı olarak 6 öğretmenin katılımıyla yürütülmüştür.

## 2.3. Veri Toplama Araçları

### 2.3.1. Performans Görevi

Performans görevi, öğrencilerin uygulama çalışmaları için kendilerine aktarılan teorik bilgileri kullanabilecekleri bir proje olarak, derse giren öğretmenler tarafından belirlenmiş ve görevin ayrıntıları tasarlanmıştır. Performans görevi ortalama bir öğrencinin yapabileceği şekilde belirlenmiştir.

### 2.3.2. Puanlama Cetveli (Puanlama Anahtarı)

Uygulamanın değerlendirilmesi için, ders öğretmenleri uygulamanın temel basamaklarını belirlemiştir. Daha sonra bu basamaklar uygulama içindeki önemlerine göre puanlanmıştır. Öğrencilerin puanlama anahtarında alacağı en düşük puan 0, en yüksek puan 40'dır. Bu çalışma 2 öğretmenle yapılmış, daha sonra uygulamaya katılacak olan diğer 4 öğretmenin de görüşleri alınarak puanlama anahtarına son şekli verilmiştir.

### 2.3.3. Analitik Dereceli Puanlama Anahtarı

Geliştirilen analitik DPA sadece bu ders için kullanılabilir özellikte ve Goodrich (2001)'in önerdiği basamaklar izlenerek hazırlanmıştır.

*1. Performansı belirlemede kullanılacak ölçütlerin listelenmesi:* Performans değerlendirmede kullanılacak olan DPA'da dersin hedefleri ve uygulama sırasında öğrencide görmek istenilen beceriler göz önüne alınarak beş ölçüt belirlenmiştir. Her bir ölçüt performansın önemli bir bölümüne odaklanmaktadır.

*2. Puanlama stratejisi olarak hangi DPA türünün kullanılacağına karar verilmesi:* DPA'nın türüne karar verilirken değerlendirmenin amacı göz önüne alınmıştır. Böyle bir uygulamayı değerlendirirken, amaç sadece çalışma sonunda elde edilecek olan bir ürün değil, öğrencinin uygulama süreci boyunca geçtiği basamaklardaki becerileri kazanıp kazanmadığını da değerlendirmektir. Bu bağlamda analitik DPA geliştirilmiştir.

*3. Performans düzeylerinin belirlenmesi ve düzey tanımlarının yapılması:* Bu aşamada, performans düzeyleri en yetkinden en zayıfa doğru belirlenmiş ve puanlanmıştır. Performans düzeyleri belirlenirken göreceli ifadelerden olabildiğince kaçınılmıştır. Uygulamanın bir avantajı olarak, iki ölçüt dışındaki tüm ölçütlerin performans düzeylerini sayılarla ifade etmek mümkün olmuştur. Bunun da puanlayıcı yanlılığını oldukça azaltacağı düşünülmektedir. Puanlamaya göre, yapılacak performans tam olarak gösteren öğrenci 4 puan, en zayıf düzeyde gösteren öğrenci ise 1 puan almaktadır.

DPA geliştirme aşamasında, alan uzmanları bazı ölçütlerin diğerlerinden daha önemli olduğunu belirtmişler ve bu ölçütler x3 ve x4 ifade edilmiş, yani 3 katı ve 4 katı ile puanlanmasına karar verilmiştir. Buna göre, en zayıf öğrenci 10 puan alırken en iyi öğrenci 40 puan almıştır.

*4. Uzman görüşünün alınması:* DPA, alan uzmanı 2 öğretmenin yönergeleri doğrultusunda hazırlanmıştır. Uygulamaya katılacak diğer 4 öğretmenin de görüşleri alınarak DPA'ya son şekli verilmiştir. Tüm bu aşamalarda ölçme ve değerlendirme uzmanlarının önerileri doğrultusunda gerekli düzeltmeler yapılmıştır.

## 2.4. Verilerin Toplanması

Araştırmada üç farklı veri toplanmıştır:

İlk veri, yapılan uygulamanın değerlendirilmesinde kullanılan puanlama anahtarından elde edilen puanlardır. Öğrencilere uygulama sınavı ile ilgili yönergeler açıklandıktan sonra süreç başlatılmıştır. Öğretmenler puanlama anahtarındaki işlem basamaklarını kontrol ederek puanlamayı tamamlamıştır.

İkinci veri, aynı uygulamanın değerlendirilmesinde kullanılan DPA'dan elde edilen puanlardır. Uygulamadan önce öğrencilerin her birine DPA dağıtılarak, ölçme aracının amacı vurgulanmış ve nasıl puanlama yapılacağı anlatılmıştır. Bu sırada, öğrencilerin tanımlanmış ölçütler doğrultusunda hangi düzeyde oldukları belirlenmiş ve performansları puanlanmıştır. Değerlendirme sonlandıktan sonra, DPA öğrencilere geri verilmiş ve her bir ölçütte nasıl değerlendirildikleri ile ilgili geri bildirim almaları sağlanmıştır.

Üçüncü veri, uygulama yapılan derse ait 1.dönem notlarıdır. Puanlama anahtarı ve DPA ile verilen puanların, ilgili dersin 1.dönem notlarını yordama düzeyine bakmak için, okulun bilgi işlem kayıtlarından sağlanmıştır.

## 2.5. Verilerin Analiz

Veriler alt problemler temel alınarak çözümlenmiştir. Birinci alt problemde, iki farklı ölçme araçları ile verilen puanlar arasındaki uyum düzeyini araştırmak için Pearson Korelasyon Katsayısına bakılmıştır. İkinci alt problemde, ilgili puanların ortalamaları arasında manidar bir fark olup olmadığını araştırmak için mutlak başarı yüzdelerine ilişkin ilişkili örneklemelerde t-testi yapılmıştır. Daha sonra, iki farklı ölçme aracı ile verilen puanlar ile 1.dönem notları arasında bir ilişki olup olmadığı araştırılmıştır. Elde edilen bulgulara göre, DPA ile verilen puanların 1.dönem notlarını yordama düzeyinin daha yüksek çıktığı görülmüştür.

## 3. BULGULAR

Bu bölümde araştırmanın problemlerine cevap bulmak için öğrenci puanlarına ilişkin toplanan tüm veriler, alt problemlerin kapsamına uygun olarak yapılan analizler, bu analizlerden yola çıkarak elde edilen bulgular ve bulgulara ilişkin yorumlar yer almaktadır.

### 3.1. Puanlama anahtarı ve DPA ile verilen puanların uyum düzeyi nedir?"

Öğrencilere iki ölçme aracı ile verilen puanlar karşılaştırılmıştır. Puanlama anahtarı ile değerlendirildiğinde ortalamasının altında kalan öğrencilerin DPA'dan 4-7 puan daha fazla aldığı ve ortalamaya yaklaştıkları görülmüştür. Ancak puanlama anahtarı ile değerlendirildiğinde, ortalamasının üstünde yer alan öğrencilerin DPA'dan da yüksek puan aldıkları görülmüştür. Bu durum, DPA ile değerlendirmenin alt gruptaki öğrencilerin lehine olduğu, üst gruptaki öğrencilerin başarı durumlarını çok etkilemediği kanısını desteklemektedir.

İki ölçme aracı ile verilen puanların uyum düzeyini belirlemek amacıyla Pearson Korelasyon Katsayılarına bakılmıştır. Korelasyonlar 0,870 ile 1,00 arasında değişmektedir. Bu bulgulara göre, her grupta farklı iki ölçme aracıyla verilen puanlar arasında yüksek bir ilişki olduğu görülmektedir.

### 3.2. Puanlama anahtarı ve DPA ile elde edilen puanların ortalamaları arasında manidar bir fark var mıdır?

Ölçme araçlarından elde edilen puanların minimum değerleri birbirinden farklı olduğu için aynı örneklem üzerinden elde edilen puanların mutlak başarı yüzdeleri alınmıştır. Bu puanların ortalamaları arasında anlamlı bir fark olup olmadığını belirlemek için bağımlı gruplarda t-testi yapılmıştır. Elde edilen sonuçlara göre, ortalamalar arasındaki farkın DPA lehine çıktığı görülmüştür (Tablo 1).

**Tablo 1: Puanlama Anahtarı ve DPA Ortalama Puanları t-Testi Sonuçları**

Puanlama Tekniği	N	$\bar{X}$	S	Sd	t	p
Puanlama Anahtarı	70	64,392	23,201	69	-6,708	,000
Dereceli Puanlama Anahtarı	70	71,035	18,787			

Birinci alt problemde, iki ölçme aracı ile verilen puanlar arasındaki korelasyon katsayılarının yüksek çıktığı göz önüne alındığında, DPA ile verilen puanların ortalamasının daha yüksek olmasının önemli olduğu söylenebilir. DPA'da, ölçütlerin ve performans düzeylerinin daha net ifade edilmesi ve dolayısıyla daha ayrıntılı bir puanlamaya imkan vermesi bu bulgunun temel nedeni olarak gösterilebilir.

### 3.3. Puanlama anahtarı ve DPA'dan elde edilen puanlar ile öğrencilerin 1.dönem notları arasındaki korelasyonların farkı manidar mıdır?

Araştırma sırasında kullanılan iki ölçme aracı ile 1.dönem notları arasında bir ilişki olup olmadığını belirlemek amacıyla oluşturulan bu alt probleme cevap aramak için, puanlama anahtarı ve DPA'dan elde edilen puanlar ile öğrencilerin 1.dönem notları arasındaki ilişkiye Pearson Korelasyon Katsayısı ile bakılmıştır. Analiz sonucunda elde edilen korelasyon katsayıları Tablo 2'te verilmiştir.

**Tablo 2: Ölçme Araçlarından Elde Edilen Puanlar ile 1.Dönem Notları Arasındaki Korelasyon Katsayısı Sonuçları**

	N	Puanlama Anahtarı	Dereceli Puanlama Anahtarı	Zr
1.Dönem Notu	70	,366(**)	,419(**)	0,38 (p>0,05)

\*\*p<0,01

Her iki korelasyon katsayısı da 1.dönem notları ile orta düzeyde bir ilişki olduğunu göstermektedir. DPA ve ders notları arasındaki korelasyon, puanlama anahtarı ile elde edilen korelasyondan biraz daha yüksek çıkmıştır. Bu bulgu DPA ile elde edilen puanların, 1.dönem sonunda verilen ders notları ile daha uyumlu olduğunu göstermektedir.

Analiz sonuçları, her iki ölçme aracıyla elde edilen korelasyon katsayıları arasında göreceli olarak bir fark olduğunu göstermektedir. Bu farkın manidar olup olmadığını test etmek için Fisher'in z testinden yararlanılmıştır, z değeri 0,38 çıkmıştır. Buna göre, iki korelasyon arasındaki farkın manidar olmadığı söylenebilir.

### 3.4. Puanlama anahtarı ve DPA'dan elde edilen puanların, 1.dönem notlarını yordama düzeyi nedir?

1.dönem sonu notları ile ölçme araçlarından elde edilen puanların arasındaki ilişkiyi araştırmak için çoklu regresyon analizinden yararlanılmıştır. Böylece hangi ölçme aracından elde edilen puanların 1.dönem notlarını daha iyi yordadığı sorusunun cevabı bulunacaktır. Çoklu regresyon analizi sonuçları Tablo 3'de verilmiştir.

**Tablo 3: Öğrenci Notlarının Yordanmasına İlişkin Çoklu Regresyon Analizi Sonuçları**

Model	Standartlaştırılmamış Katsayılar		Standartlaştırılmış Katsayılar	t	P	R	R <sup>2</sup>
	B	Standart hata	Beta				
Sabit	32,417	7,662		4,231	,000		
Puanlama Anahtarı	-,179	,218	-,275	-,824	,413	,429	,184
Dereceli Puanlama Anahtarı	,547	,269	,678	2,035	,046		

Regresyon denkleminin önemli olup olmadığını anlamak için ANOVA değerlerine bakılmıştır. F = 7.56 değerinin 0.05 düzeyinde manidar çıkmasına dayanarak değişkenlerden en az birinin yordama düzeyinin manidar olduğuna karar verilmiştir. Öğrenci notlarının yordama katsayılarına ilişkin t değerlerine bakıldığında, DPA ile verilen puanlarının yordama düzeyinin 0,05 düzeyinde istatistiksel olarak manidar olduğu görülürken, puanlama anahtarına ilişkin

puanların yordama düzeyinin manidar olmadığı görülmektedir. Tablo 3'e bakıldığında çoklu korelasyonun  $R=0.43$  ve çoklu korelasyonun karesinin (açıklama katsayısı)  $0,18$  olduğu görülmektedir. Buna göre, puanlama anahtarı ve DPA'nın birlikte okul notlarındaki değişkenliğin %18'ini açıkladığı söylenebilir. Bu değere dayanarak, puanlama anahtarı ve DPA ile elde edilen puanların öğrenci notlarını önem verilecek düzeyde açıkladığı yorumu yapılabilir. Standardize edilmiş regresyon katsayısı beta değerine göre, okul notları üzerinde DPA'nın daha önemli bir yordayıcı olduğu söylenebilir.

#### 4. TARTIŞMA VE SONUÇ

Ölçme araçları ile verilen puanların karşılaştırıldığında, DPA ile verilen puanların öğrencinin lehine olduğu görülmektedir. Bu durum, öğrencilerin değerlendirme ölçütleri hakkında önceden bilgi sahibi olmalarının ve kendilerinden ne beklediğini bilmelerinin performans üzerinde olumlu etkileri olduğu şeklinde yorumlanabilir. Ayrıca, DPA ile verilen puanlar ile öğrencilerin 1.dönem notları arasındaki ilişkiye bakıldığında, elde edilen korelasyon katsayısı ve yordama düzeyi, puanlama anahtarından elde edilen değerlerden göreceli olarak daha yüksektir.

Öğrencilere puanlama anahtarı ve DPA ile verilen puanlar tek tek karşılaştırıldığında, ortalamanın altında yer alan öğrencilerin, puanlama anahtarına göre DPA'dan daha yüksek olarak ortalamaya yaklaştıkları, ortalamanın üstünde yer alan öğrencilerin ise iki ölçme aracı ile aldıkları puanların birbirine çok yakın olduğu tespit edilmiştir. Bu durum, DPA ile değerlendirmenin alt gruptaki öğrencilerin lehine olduğu, üst gruptaki öğrencilerin başarı durumlarını çok etkilemediği kanısını desteklemektedir. Özellikle Analitik DPA'da ara basamaklar ayrıntılı puanlandığı için, performansı düşük olan öğrencilerin daha avantajlı olduğu söylenebilir. Çoktan seçmeli ve kısa yanıtli testlerde gerçek başarılarını sergileyemeyen öğrenciler için DPA ile değerlendirmenin alternatif bir fırsat yarattığını ifade edilmektedir (Airasan ve Russel 2008).

Elde edilen bulgular ve ulaşılan sonuçlara dayanarak okullardaki uygulamalar sırasında, değerlendirmenin duyarlı ve objektif yapılabilmesi ve sonuçlarının doğru yorumlanabilmesi için, DPA'nın yapısı ve puanlamasıyla ilgili öğretmenlere daha ayrıntılı bilgi verilmeli ve çalışmanın amacı daha iyi anlatılmalıdır. DPA'nın ayrıntılı yapısından sıkılan ya da alışkanlıklarından dolayı önceki puanlama yöntemlerine göre puanlama yapan puanlayıcılar olabilmektedir. Bundan dolayı elde edilen veriler sağlıklı sonuçlar almayı ve yorumlar yapmayı zorlaştırabilir. Yapılan uygulamalarda, aynı ölçme aracı ve aynı öğrencilerle farklı öğretmenlerin uygulama yapabilme olanağı yaratılırsa, yapılan değerlendirmenin geçerliği ve güvenilirliği üzerinde daha genellenebilir sonuçlar elde edilecektir. Ayrıca bu şekilde, aynı puanlayıcıların farklı puanlama yaptıkları ya da yanlış davrandıkları ölçütler tespit edilebilir ve ölçme aracı sonraki uygulamalar için düzeltilerek daha iyi sonuçlar elde edilebilir.

#### 5. KAYNAKLAR

- Airasian, P. W. & Russel, M. K. (2008). *Classroom assessment. Concepts applications*. USA: McGraw-Hill.
- Akhun, İ. (1992). İki Korelasyon Arasındaki Farkın Manidarlığının Test Edilmesi.
- Aslanoğlu, E. A.ve Kutlu, Ö. (2004). Eğitimde sunu becerilerinin değerlendirilmesinde dereceli puanlama anahtarı uygulamasına ilişkin bir araştırma. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, cilt 36, sayı: 1-2.
- Bekiroglu, F. (2004). *Ne kadar başarılı? Klasik ve alternatif ölçme değerlendirme yöntemleri ve fizikte uygulamalar*, Ankara: Nobel Yayın.
- Biemer, L. (1993). Authentic assessment. *Educational leadership*, 50 (8), 81-82.
- Brookhart, S. M. (1999). The art and science of classroom assessment: The missing part of pedagogy. ASHE-ERIC.
- Büyüköztürk, Ş. (2007). Performansa dayalı durum belirleme nedir? *İlköğretmen Eğitimci Dergisi*, Sayı 8,28-32.

- Callison, D. (2000). Rubrics. *School Library Media Activities Monthly*, 17 (2).
- Goodrich, H. (1997). Understanding rubrics. *Educational Leadership*, 54(4), 14-17.
- Goodrich, A. H. (2001). The effects of instructional rubrics on learning to write. *Current Issues ID Education*.
- Goodrich, A. H. (2005). Teaching with rubrics: the good, the bad and the ugly. *Current Issues in Education*, 4 (4).
- Gündüz S., D. (2006). *Matematik dersinde problem çözme becerilerinin dereceli puanlama anahtarı kullanılarak değerlendirilmesi*, Yayınlanmamış yüksek lisans tezi, Hacettepe Üniversitesi, Ankara.
- Haladyna, T. M. (1997). *Writing test item to evaluate higher order thinking*. USA: Allyn & Bacon.
- Higher Education Report (Vol. 27, 69 No.1). Washington, DC: The George Washington University, Graduate School of Education and Human Development.
- Kutlu, Ö., Doğan, C., D. ve Karakaya, İ. (2008). Öğrenci Başarısının Belirlenmesi. Ankara: PegemA Yayıncılık.
- McMillan, J. H. (1997). *Classroom assessment: principles and practice for effective instruction*. Needham, MA: Allyn & Bacon.
- Mehrens, W. A. (1992). *Using performance assessment for accountability purposes*. *Educational Measurement*, 11, (4) 3-9.
- Moskal, B. M. (2000). "Scoring rubrics: What, when and how?" *Practical assessment, Research & Evaluation*, 7 (3).
- Popham, J. W. (1997). *What's wrong and what's right with rubric*. *Educational Leadership*, 55, (2), 12
- Popham, J. W. (2000). *Modern Education Measurement*. Needham: Allyn & Bacon.
- Sezer, S. öğrencinin akademik başarısının belirlenmesinde tamamlayıcı değerlendirme aracı olarak rubrik kullanımı üzerinde bir araştırma. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, Sayı:18
- Sezgin, S. İ. (2000). *Mesleki ve teknik eğitimde program geliştirme*. Ankara: Nobel Yayın Dağıtım.
- Stiggins, R. J. (1994). *Student-Centered classroom assessment*. New York: Macmillan Publishing Company.
- Şenel, T., Yıldırım, N. ve Er Nas, S. (2007). *Süreç odaklı değerlendirmede kullanılacak bir analitik rubriğin geliştirilmesi: yaşamımızdaki elektrik ünitesi örneği*. *Edu 7*, Cilt 2, Sayı 2, Eylül 2007
- Tekin, H. (1991). *Eğitimde ölçme ve değerlendirme*, Ankara: Yargı Yayınları
- Turgut, M. F. (1997). *Eğitimde ölçme ve değerlendirme Metotları*. Ankara: Yargıcı Matbaası.

### Extended Abstract

Radical and rapid changes that affect the society entail changes in education as in many areas. As we will be up against rapidly increasing information in the new era, individuals who know how and where to get information, who are selective, and who learn how to learn, rather than individuals who know everything will be demanded. While the information keeps renewing rapidly, it has also radically changed professional, technical, educational knowledge and skills expected from individuals. It is obvious that processes of producing quality education products, final products and the adequacy of approaches to select the method to evaluate the performances will affect the success of the individual.

Under these developments, it has become necessary to make changes curriculum in the education process, methods and techniques, assessment and evaluation approaches. One of the different evaluation methods we see here is the evaluation of performance. Defines the evaluation of performance as a method where the student is asked to generate an answer or show his/her knowledge by creating a product rather than asking student to find a single correct answer.

Mostly, performance in education is evaluated indirectly. However, this is inconvenient for the validity of the evaluation results. Evaluating the knowledge, which indicates the performance, is not the proof of how that performance can be done. It is because there is not always a direct correlation between knowing how to do it and doing it according to a set standard. Therefore, we need to use proper evaluation instruments to assess the performance. Especially in subjects such as industrial crafts and laboratory application where there are relevant goals for performance it is quite difficult to conduct assessment and evaluation by means of pen and paper tests, and it involves a validity issue. It is because in a performance evaluation, the student is asked to do things rather than to explain how to do it. Scoring rubric is one of the



most commonly used instruments for evaluating performance. Scoring rubric minimizes the errors that may occur during scoring with its detailed structure consisting of standards and performance levels.

The research was conducted in the electric & electronics major of a vocational high school on 70 students and 6 teachers. At first the performance task was defined for the research. The performance task was determined as a project where students can use given theoretical information for the application, and the details for the task were designed. The answer key and scoring rubric for the evaluation of the application were prepared later. While the answer key is prepared, the course teachers set the basic stages of the application and rate the stages based on their importance and the answer key took its final form. The analytic scoring rubric developed can be used only in this course and it is prepared by following the steps suggested. The final standards and performance levels were formed as per the studies of the 6 teachers in the research.

In this research we aim to determine the benefits of using answer key and scoring rubric during the evaluation for the courses that requires application for the learning process and to research the adjustment levels of scorings of two different evaluation instruments by comparing evaluation scores. When comparing the scorings of two different evaluation instruments, the adjustment levels of these scores and the course scores were reviewed. The t test was conducted to determine the relation between the average scores obtained by using answer key and scoring rubric and it is seen that the average score obtained by using scoring rubric are higher. The correlation between the points for the first semester and the points obtained by using both scoring instruments were examined and it was seen that the correlation of the points obtained by scoring rubric is higher. However, the coefficient difference between two correlations was 0.05 and it is not significant. As a result of regression analyses, the regression level of points obtained by scoring rubric was higher than that of school points.

These are the results based on the findings obtained. When we compare the points obtained by using answer key and scoring rubric one by one, we see that students who get under average points based on answer key get higher points based on scoring rubric and come close to the average. On the other hand, the points were very close to each other for students who get above average or higher points based on both scoring instruments. Therefore, this supports the assumption that evaluation by scoring rubric is in favor of the students with lower points and has no affect on the success of students with higher points. Especially in Analytical scoring rubric since the intermediary levels are scored in detailed, students with low performance have an advantage. Evaluation by scoring rubric creates an alternative opportunity for the students who do not demonstrate their real performances in multiple choice or short answer tests. Another finding is that the correlation between the points for the first semester and the points obtained by using scoring rubric were examined and the correlation coefficient and regression level obtained was higher than the values obtained by using answer key.

---

## Kaynakça Bilgisi

Parlak, B. ve Doğan, N. (2014). Dereceli puanlama anahtarı ve puanlama anahtarından elde edilen puanların uyum düzeyleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi [Hacettepe University Journal of Education]*, 29(2), 189-197.

## Citation Information

Parlak, B., & Doğan, N. (2014). Comparison of answer key and scoring rubric for the evaluation of the student performances. [in Turkish]. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi [Hacettepe University Journal of Education]*, 29(2), 189-197.