

PIYANO EĞİTİMİ DERSİ PERFORMANS ÖLÇÜMÜNE YÖNELİK DERECELİ PUANLAMA ANAHTARININ GELİŞTİRİLMESİ: GEÇERLİK VE GÜVENİRLİK ÇALIŞMASI

Araş. Gör. Şafak TURGUT*
Prof. Dr. Esra DALKIRAN**

Özet: Çalgı performans sınavlarında, likert tipi ölçekler ve analitik DPA türü ölçme araçları kullanıldığında, hakemler öğrenci performansını dinlemek, ölçme aracı maddelerini okuyup anlamak ve performansa göre uygun puanlamayı yapmak durumundadır. Dolayısıyla araştırmacılar, bu ölçme araçlarının çalgı sınavı gibi kısıtlı bir zaman içerisinde doldurmaya uygun olmadığını düşünmektedir. Araştırmanın amacı, piyano eğitimi dersi sınav performansını ölçmeye yönelik geçerliği ve güvenirliliği sağlanmış, sürecin bütününe odaklanan, pratik bir ölçme aracı geliştirilmesidir. Tarama modeli kullanılan araştırmanın çalışma grubunu, 2018-2019 eğitim-öğretim yılında, üniversitelerin, eğitim fakültelerine bağlı müzik eğitimi anabilim dalında, lisans 1 düzeyinde öğrenim gören 52 müzik öğretmeni adayı oluşturmaktadır. Araştırmanın verileri, araştırmacılar tarafından geliştirilen DPA ile toplanmış, sınıf içi korelasyon analizi (Intraclass Correlation Coefficient) ile genel puanlar ve tüm ölçütlere verilen puanlar arasında hakemler arası tutarlılığa bakılmıştır. DPA'nın güvenirlilik analizleri ve iç tutarlılığının belirlenmesinde Cronbach's alpha ile madde-bütün analizleri (Item-Total Statistics) yapılmıştır. DPA'nın kapsam geçerliği için uzman görüşü alınmış; yapı geçerliği için ise, iki farklı ölçme yöntemi arasındaki ilişkinin belirlenmesi amacıyla Pearson korelasyon analizi kullanılmıştır. Araştırmanın sonuçlarına göre, hakemler arası tutarlılık, .850 ve .976 arasında değişmektedir. Tüm ölçütler üzerinden hesaplanan Cronbach's alpha iç tutarlık katsayısı, hakem-1 ve hakem-2 için .80'den büyük olduğundan, DPA'nın güvenirliliği yüksek düzeydedir. DPA'nın, yapı geçerliği için yapılan Pearson korelasyon analizi sonuçlarına göre; hakem-1 ve hakem-2 için ölçme yöntemleri arasında yüksek korelasyon bulunmuştur.

Anahtar Kelimeler: Müzik, Piyano, Performans, Dereceli Puanlama Anahtarı, Ölçme Ve Değerlendirme

Geliş Tarihi: 16.12.2020

Kabul Tarihi: 28.09.2021

Makale Türü: Araştırma Makalesi

*Kastamonu Üniversitesi, Eğitim Fakültesi, Güzel Sanatlar Eğitimi Bölümü, sturgut@kastamonu.edu.tr, ORCID: 0000-0002-0279-6407

**Burdur Mehmet Akif Ersoy Üniversitesi, Eğitim Fakültesi, Güzel Sanatlar Eğitimi Bölümü, edalkiran15@gmail.com

ORCID: 0000-0001-6829-7216

DEVELOPMENT OF A PIANO EDUCATION LESSON PERFORMANCE ASSESSMENT RUBRIC: THE STUDY OF VALIDITY AND RELIABILITY

Res. Assist. Şafak TURGUT*
Prof. Dr. Esra DALKIRAN**

Abstract: When Likert-type scales and analytical rubric type measurement tools are used in instrument performance exams, the adjudicators have to listen the student performance, read and understand the measurement tools and make appropriate scoring according to the performance. Therefore, the researchers think that these measurement tools are not suitable for filling in a limited time like an instrument exam. The aim of the research is to develop a practical measurement tool that focuses on the whole process, whose validity and reliability have been ensured for the piano education lesson. Survey model was used in this study. The study group of the research consists of 52 pre-service music teachers, studying at undergraduate level 1 in the department of music education. The data of the study was collected with rubric developed by the researchers. Intraclass Correlation Coefficient was used to determining inter-rater reliability for general scores and scores given to all criteria. Cronbach's alpha and item-total statistics were used to determine the reliability and internal consistency of the rubric. Expert opinion was obtained for the content validity of rubric. For the construct validity of the rubric, Pearson correlation analysis was used to determine the relationship between two different assessment methods. According to the results of the study, the consistency between the adjudicators varies between .850 and .976. Since the Cronbach's alpha coefficient calculated over all criteria is greater than .80 for adjudicator-1 and adjudicator-2, the reliability of the rubric is high. According to the results of Pearson correlation analysis; There is a high correlation between assessment methods for adjudicator-1 and adjudicator-2.

Keywords: Music, Piano, Performance, Rubric, Assessment

Received Date: 16.12.2020 Accepted Date: 28.09.2021 Article Types: Resarch Article

*Kastamonu University, Faculty of Education, Fine Arts Education Department, sturgut@kastamonu.edu.tr, ORCID: 0000-0002-0279-6407

**Burdur Mehmet Akif Ersoy University, Faculty of Education, Fine Arts Education Department, edalkiran15@gmail.com, ORCID: 0000-0001-6829-7216

1. GİRİŞ

Sanat alanında ölçme ve değerlendirmenin objektif ve adil koşullarda yapılması önem arz etmektedir. Birçok ders için sorular ve bu soruların kesin cevapları bulunmaktadır ancak müzik eğitiminde performans değerlendirmesi doğası gereği subjektif bir değerlendirme değildir. Geçerliliği ve güvenilirliği sağlanmış ölçme araçları bu subjektifliği ortadan kaldırmada önemli yer tutmaktadır. Kırıçoğlu'na (2005) göre; "sanat, okul sistemleriyle bağdaşsa da bağdaşmasa da bu kurumlar içinde yer aldığı sürece not ile ölçme ve değerlendirme kaçınılmazdır. Ayrıca bu derste derecelendirmeden kaçınmak derse yaklaşımlardaki olumsuz tutumu da artıracaktır". "Ölçme, bireylerin ya da nesnelere belirli özelliklere sahip olup olmadığının, sahipse, sahip oluş derecesinin belirlenerek sonuçların sembollerle, özellikle sayı sembollerıyla ifade edilmesidir. Değerlendirme ise, kıstaslar kullanarak ölçme sonuçlarına göre nitelik hakkında bir karara varma sürecidir" [Millî Eğitim Bakanlığı (MEB), 2007].

"Değerlendirme işleminin yapılabilmesi için ona dayanak noktası oluşturacak ölçme sonuçlarına ihtiyaç vardır. Ölçme işlemi ve ölçme işleminden elde edilen ölçme sonuçları olmadan değerlendirmenin yapılması olanaksızdır. Ancak ölçme sonuçlarının sahip olduğu geçerlik ve güvenilirlik derecesi, değerlendirmeyi olumlu veya olumsuz yönde etkileyebilmektedir" (Yaşar, 2008). Sanat eğitiminde ölçme ve değerlendirme süreci diğer alanlara kıyasla, kesin sonuçlara varmak açısından daha problemlidir. Cornett ve Smithrim'e (2001) göre, örgün eğitim içerisindeki sanat eğitimi ile ölçme ve değerlendirme arasındaki gerilim, geçmişte ve günümüzde bir eğitim ikilemi oluşturmaktadır. "Eğitim süreci içerisinde ölçme ve değerlendirmenin ne derece gerekli olduğu ve sağlam temellere oturması gerektiği bilinse de sanat eğitiminin diğer alanlara kıyasla daha karmaşık yapısından dolayı,

bu alanda ölçme ve değerlendirme son derece subjektif ölçümlerle yapılmaktadır" (Tuna, 2011).

"Sanat eğitiminde çalışma alanları içinde en zor değerlendirilen ise, uygulamalı çalışmalarda üründür. Öğrenciden duygu, düşünüyü ve imgelerini sanatsal niteliği olan bir forma veya performansa dönüştürmesi istendiğinde, sonucun değerlendirilmesi de beklenir" (Kırıçoğlu, 2005: 207). Müzik eğitiminde uygulamalı çalışmalardan biri olan çalgı eğitimi, müzik eğitiminin vazgeçilmez bir parçasıdır. Dolayısıyla çalgı eğitimi dersleri, Güzel Sanatlar Liselerinin ve Üniversitelerin müzik bölümlerinin/anabilim dallarının programlarında da yerini almıştır. Uçan'a (1980) göre, "çalgı eğitimi; çalgı öğretimi yoluyla bireylerin ve onların oluşturdukları toplulukların devinimsel, bilişsel ve duyuşsal davranışlarında kendi yaşantıları yoluyla ve kasıtlı olarak istedik değişiklikler oluşturma, bu nitelikte yeni davranışlar kazandırma sürecidir".

Öğrencinin, çalgı performansının değerlendirilmesi müzik eğitimcileri için her zaman zor bir süreç olmuştur. Sanat eğitimi ve müzik eğitimi alanında birçok ders performansın değerlendirilmesine bağlıdır. Performansın değerlendirilmesi genellikle öznel bir değerlendirme olarak görülmektedir. Radocy'e (1989) göre, İnsani yargılar içeren herhangi bir ölçme süreci, doğası gereği öznel, çünkü insan izlenimleri içermektedir. Ancak, Radocy farklı bir araştırmada şunu da belirtmektedir: Yalnızca insan gözlemciler bir sesin veya ses dizisinin müzikal değerini başarılı bir şekilde değerlendirebilir (Radocy, 1986).

Yapılan incelemeler sonucunda, literatürde çalgı performans sınavında doldurulabilecek, pratik ölçme aracı eksikliği olduğu düşünülmektedir. Performans sınavlarında her öğrenci için ayrı sınav yapıldığından, öğrenciye ayrılacak zaman kısıtlıdır. Hakemler, genelde 4-5 dk süren performans sınavında, ölçme aracının

maddelerini okumak, işaretlemek ve aynı zamanda öğrenciyi dinlemek durumunda kalmaktadır. Bu durum hakemleri, sınavlarda ölçme aracı kullanmamaya itebilmektedir. Ölçme araçları, her alan için olduğu kadar müzik alanı için de gereklidir.

Amerika Ulusal Müzik Eğitimi Derneği'ne (MENC, 1996) göre, güvenilir bir müzik ölçme aracı şu özellikleri taşımalıdır: (a) aynı beceri veya bilginin ölçülmesi aynı sonuçları vermelidir, (b) her performans aynı prosedür takip edilerek değerlendirilmelidir ve (c) hakemlerin aynı düzeyde beklentileri olmalıdır. Müzik alanında, akademik ve performansa yönelik olarak dereceli puanlama anahtarı (DPA) geniş ölçüde kullanılmaya başlanmıştır. DPA müzik öğrencilerine hedefe yönelik daha fazla bilgi vermesiyle ve DPA dışındaki ölçme araçlarına göre öğrencinin gelişimine yönelik daha kesin sonuçlar ortaya koymasıyla tercih edilmektedir.

Kan'a (2007) göre; "DPA, yapılandırılmış performans görevleri üzerinde değişik düzeylerde performansa ait karakteristik özellikleri ve ölçütleri tanımlayan, bu özellik ve kriterler doğrultusunda performansa ya da ürüne ilişkin yargıya varmada kullanılan puanlama rehberidir". Goodrich (2005)'e göre; eğitim sürecinde dereceli puanlama anahtarı kullanımının yararları şunlardır: "(a) öğretmen, öğrenci, veli ve başka değerlendiriciler için ölçütlerin açık olmasını sağlar, (b) öğrencilerden neler beklendiğini açıkça gösterir, (c) durum belirlemede nesnellik sağlar, (d) öğrenciye verilen notların doğru bir biçimde açıklanmasını sağlar, (e) öğrencilerin güçlü ve zayıf yönleri konusunda bilgilendirici geri bildirim verir, (f) öğretim sürecinin gelişmesine katkı sağlar" (aktaran: Kutlu, Doğan ve Karakaya, 2017: 74).

DPA, öğretmenin derse yönelik beklentilerini açıklayarak, öğrencilere performansa daha düzenli ve programlı hazırlanma imkânı sunmaktadır. "Öğrencilere verilen performans

görevinde belirli özellik ve düzeylere göre dikkat edilmesi gereken nitelikleri tanımlayan bir ölçme aracı olduğundan, performans sonucu ortaya çıkan ürünün mükemmel niteliklerinin ne olması gerektiği konusunda da bilgi verir" (Arı, 2008).

DPA'yı diğer ölçme araçlarından ayıran özelliklerden bazıları şunlardır: "Tahminde bulunma, geribildirim sağlama ve not verme özelliğinin olması, DPA'yı diğer ölçme araçlarından ayırmaktadır" (Moskal, 2000). Saddler ve Andrade'ye (2004) göre, "DPA planlama, gözlemlene ve düzenleme gibi üstbilgi stratejilerini destekleyerek öğrencilerin öz düzenleme becerilerini geliştirir. Öğrencinin öğrenmesini, eksikliklerini gidermesini sağlayan DPA, üst düzey düşünme becerileri ile üstbilgi becerilerinin birlikte gelişimini destekler". Arter'e (2000) göre "Üst düzey düşünme becerilerine odaklanan ve bu becerilerin gözlenmesine olanak sağlayan yeni değerlendirme yaklaşımlarından biri olan DPA, değerlendirilecek ürün ya da performansla ilgili olarak öğrencilerin üst düzey düşünme becerilerinin sürekli olarak artmasına katkıda bulunur".

Likert tipi ölçekler, birçok çalışmada solo müzik performansını değerlendirmek için kullanılmıştır (Abeles, 1973; Bergee, 1988, 1989; Grançer Okay, 2010; Horowitz, 1994; Jones, 1986; Kurtuldu, 2010; Polat, 2017; Yazar, 2010, Zdzinski ve Barnes, 2002). Ancak, Saunders ve Holahan'a (1997) göre likert tipi ölçekler, performansın başarılı veya başarısız olmasına neyin sebep olduğu konusunda çok az bilgi içermektedir. Çünkü, likert tipi ölçekler genel ifadelere verilen puanlamalardan ibarettir, her ölçütün farklı başarı seviyelerine ilişkin açıklamalar içermemektedir.

Müzik eğitimi alanında performans değerlendirmesinde, geçerliği ve güvenilirliği sağlanmış DPA kullanılması ve ölçme değerlendirme sonucunun DPA ile birlikte, öğrencilere sunulması; öğrencilerin derse yönelik

mevcut durumlarını anlamları ve geliştirmeleri gereken davranışların farkına varmaları açısından oldukça önemlidir. Bu araştırmada, geçerliği ve güvenilirliği sağlanmış, sınav anında doldurulabilecek pratik bir DPA geliştirilmesi amaçlanmaktadır. Bu amaç doğrultusunda aşağıdaki sorulara yanıt aranmıştır:

1. Ölçüt puanlarında ve genel puanlarda hakemler arası tutarlılık hangi düzeydedir?
2. Ölçüt puanlarının iç tutarlılığı hangi düzeydedir?
3. Genel izlenim yöntemiyle yapılan ölçümlerle, DPA kullanılarak yapılan ölçümler arası korelasyon hangi düzeydedir?

2. YÖNTEM

Araştırmada, tarama modeli kullanılmıştır. Karasar (2008)'a göre "tarama modeli, geçmişte ya da halen var olan bir durumu var olduğu şekliyle betimlemeyi amaçlayan araştırma yaklaşımıdır. Araştırmaya konu olan olay, birey ya da nesne, kendi koşulları içinde ve olduğu gibi tanımlanmaya çalışılır".

Araştırmanın çalışma grubunu, 2018-2019 eğitim-öğretim yılında, üniversitelerin, eğitim fakültelerine bağlı müzik eğitimi anabilim dalında, lisans 1 düzeyinde öğrenim gören 52 müzik öğretmeni adayını oluşturmuştur. Müzik öğretmeni adayları, piyano eğitimi dersi final sınavı esnasında iki hakem tarafından değerlendirilmiştir. Araştırmada, çalışma gurubunu oluşturan 52 müzik öğretmeni adayının piyano performansları, piyano eğitimi alanında uzman iki hakem tarafından hem genel izlenim yoluyla hem de DPA kullanılarak puanlanmıştır.

Veri toplama aracı olarak, araştırmacılar tarafından geliştirilen DPA kullanılmıştır. DPA geliştirilirken, müzik öğretmeni yetiştirme programında yer alan Piyano Eğitimi ders içerikleri analiz edilmiş, sonrasında literatür taraması yapılarak, müzik eğitiminde kullanılan

performans değerlendirmeye yönelik ölçekler ve dereceli puanlama anahtarları incelenmiştir (Kurtuldu, 2010; Grançer Okay 2010; Özalp ve Dalkıran, 2018). Çepni (2011) çalışmasında iki tür dereceli puanlama anahtarından (DPA) bahsetmiştir; holistik DPA ve analitik DPA. "Holistik DPA, genellikle yazma ve diğer yaratıcı çalışmaların yanında sanatsal çalışmalar için de tercih edilmektedir. Holistik DPA, ürün ya da süreci parçalara ayırmak ya da her ölçütü bireysel olarak değerlendirmek yerine ürün ya da sürecin bütününe odaklanmaktadır". Bu araştırmada, öğrenci performansının kısa sürede değerlendirilmesi gerektiği düşünüldüğünden, holistik DPA tercih edilmiştir.

DPA'nın ölçütleri belirlenirken, altı farklı üniversitenin piyano dersi öğretim elemanları ile sözlü olarak görüşülmüş ve piyano eğitimi dersi performans değerlendirmesinde nelere dikkat edildiği konusunda görüş alınmıştır. Literatür taraması ve alınan görüşler doğrultusunda taslak DPA oluşturulmuş ve uzman görüşüne hazır hale getirilmiştir (Şekil 1). Sonraki aşamada ise; müzik eğitimi alanından beş farklı öğretim elemanından alınan uzman görüşleri doğrultusunda gerekli düzeltmeler yapılarak, DPA ön uygulamaya hazır hale getirilmiştir.

Araştırmada, hakemler arası tutarlılığa, sınıf içi korelasyon analizi (Intraclass Correlation Coefficient) ile hem genel puanlar hem de tüm ölçütlere verilen puanlar arasında bakılmıştır. Portney ve Watkins'e (2000) göre sınıf içi korelasyon katsayısı .50'den düşük değerler zayıf tutarlılığı, .50 ve .75 arası değerler orta düzey tutarlılığı, .75 ve .90 arası değerler iyi düzey tutarlılığı ve 0.90 üzeri değerler mükemmel düzey tutarlılığı göstermektedir. DPA'nın güvenilirliği için, iç tutarlılığının belirlenmesinde Cronbach's alpha ve madde-bütün analizleri (item-total statistics) uygulanmıştır. DPA'nın kapsam geçerliği için uzman görüşü alınmış; yapı geçerliği içinse, iki farklı ölçme yöntemi arasındaki ilişkinin

Etüt(ler)te Çalım Teknikleri	Gerekli	Madde yapı ile ilişkili fakat gereksiz	Madde hedeflenen yapıyı ölçmez
Etüt(ler); çalım tekniklerine uygun şekilde seslendirildi.			
Etüt(ler); çoğunlukla çalım tekniklerine uygun şekilde seslendirildi.			
Etüt(ler); kısmen çalım tekniklerine uygun şekilde seslendirildi.			
Etüt(ler); nadiren çalım tekniklerine uygun şekilde seslendirildi.			
Etüt(ler)te Müzikalite	Gerekli	Madde yapı ile ilişkili fakat gereksiz	Madde hedeflenen yapıyı ölçmez
Etüt(ler); müzikal becerilere uygun şekilde seslendirildi.			
Etüt(ler); çoğunlukla müzikal becerilere uygun şekilde seslendirildi.			
Etüt(ler); kısmen müzikal becerilere uygun şekilde seslendirildi.			
Etüt(ler); nadiren müzikal becerilere uygun şekilde seslendirildi.			

Şekil 1. Hazırlanan uzman görüşü formunun bir bölümü

belirlenmesinde Pearson korelasyon analizi uygulanmıştır.

3. BULGULAR VE YORUM

Bu bölümde araştırma bulguları ve bulgulara dayalı olarak yapılan yorumlar yer almaktadır.

3.1. Hakemler Arası Tutarlılık

Hakemler arası tutarlılığın belirlenmesi amacıyla

Tablo 1'de görüldüğü üzere, genel izlenim yöntemi ve DPA kullanılarak yapılan değerlendirmeler için sınıf içi korelasyon katsayısı .900 üzerindedir. Bu değerler, hakemler arası tutarlığın mükemmel düzeyde olduğunu göstermektedir. Tüm ölçütlerde hakemler arası tutarlılık Tablo 2'de verilmektedir.

Ölçme Yöntemi	Sınıf İçi Korelasyon Katsayısı	%95 Güven Aralığı	
		Alt Sınır	Üst Sınır
Genel İzlenim	.974	.949	.986
DPA	.965	.851	.986

Tablo 1. Genel İzlenim Yoluyla ve DPA Kullanılarak Yapılan Değerlendirmelerin Hakemler Arası Tutarlılığı

Ölçütler	Sınıf İçi Korelasyon Katsayısı	%95 Güven Aralığı	
		Alt Sınır	Üst Sınır
Temel Davranışlar	.850	.738	.914
Dizi, Kadans, Arpej (DKA)	.971	.943	.984
Etütte Bütünlük ve Notasyon	.976	.927	.983
Etütte Çalım Teknikleri	.928	.851	.962
Etütte Müzikalite	.941	.895	.966
Eserde Bütünlük ve Notasyon	.964	.904	.983
Eserde Çalım Teknikleri	.938	.850	.970
Eserde Müzikalite	.928	.849	.962

Tablo 2. Tüm Ölçütler İçin Yapılan Değerlendirmelerin Hakemler Arası Tutarlılığı

hem genel izlenim yöntemi hem de DPA kullanılarak yapılan değerlendirmeler arasında tutarlılık ve tüm ölçütler (sekiz ölçüt) için hakemler arası tutarlılık incelenmiştir (Tablo 1 ve 2).

Tablo 2'de görüldüğü üzere, yedi ölçütte yapılan değerlendirmeler için sınıf içi korelasyon katsayısı .900 üzerinde, bir ölçütte ise .850 düzeyindedir. Ölçütler için hakemler arası tutarlılık, mükemmel ve iyi düzeydedir.

3.2. Ölçüt Puanlarının İç Tutarlılığı.

DPA'nın iç tutarlılığının belirlenmesi amacıyla öncelikle Cronbach's alpha analizi yapılmıştır (Tablo 3).

Tablo 5'te görüldüğü üzere, hakem-2'ye "ilişkin ölçüt güvenilirliğinde, bir ölçüt dışında DPA'nın genel güvenilirliğini düşüren ölçüt bulunmamaktadır.

Hakemler	Ölçme Aracı	Ölçüt Sayısı	Güvenilirlik Katsayısı (α)
Hakem-1	DPA	8	.824
Hakem-2			.815

Tablo 3. DPA'ya İlişkin Cronbach's Alpha Analizi Sonuçları

Ölçütler	Düzeltilmiş Madde-Bütün Korelasyonu	Madde Silinirse Cronbach's Alpha Değeri*
Temel Davranışlar	.786	.803
Dizi, Kadans, Arpej (DKA)	.585	.803
Etütte Bütünlük ve Notasyon	.620	.824
Etütte Çalım Teknikleri	.682	.793
Etütte Müzikalite	.630	.799
Eserde Bütünlük ve Notasyon	.520	.822
Eserde Çalım Teknikleri	.662	.797
Eserde Müzikalite	.646	.801

* Cronbach's Alpha: .824

Tablo 4. Hakem-1'e İlişkin Madde-Bütün Analizi Sonuçları

Ölçütler	Düzeltilmiş Madde-Bütün Korelasyonu	Madde Silinirse Cronbach's Alpha Değeri*
Temel Davranışlar	.827	.791
Dizi, Kadans, Arpej (DKA)	.629	.786
Etütte Bütünlük ve Notasyon	.530	.830
Etütte Çalım Teknikleri	.583	.791
Etütte Müzikalite	.639	.787
Eserde Bütünlük ve Notasyon	.528	.807
Eserde Çalım Teknikleri	.610	.789
Eserde Müzikalite	.673	.786

* Cronbach's Alpha: .815

Tablo 5. Hakem-2'ye İlişkin Madde-Bütün Analizi Sonuçları

Tablo 3'te görüldüğü üzere, her iki hakem için de Cronbach's alpha güvenilirlik katsayısı yüksek düzeydedir (>.800; Gliem & Gliem, 2003).

Her ölçütün güvenilirliğe etkisini belirlemek amacıyla her iki hakem için madde-bütün analizi (item-total statistics) yapılmıştır (Tablo 4 ve 5).

Tablo 4'te görüldüğü üzere, hakem-1'e ilişkin ölçüt güvenilirliğinde, DPA'nın genel güvenilirliğini düşüren ölçüt bulunmamaktadır.

Etütte Bütünlük ve Notasyon" ölçütü, genel güvenilirliği .015 düzeyinde düşürmektedir. Hakem-1'e ilişkin analizlerde güvenilirlik düşüşü görülmediğinden önemli bir anlam ifade etmemektedir. Ciorba ve Smith'in (2009) araştırmasında, bazı hakemlerin güvenilirliği kısmen düşürdüğü görülmektedir.

	n	\bar{x}	ss	DPA Hk-1	DPA Hk-2	Gen. İzl. Hk-1	Gen. İzl. Hk-2
DPA Hk-1		45,88	16,21	1	.957	.960	.933
DPA Hk-2	52	49,60	15,91		1	.940	.959
Gen. İzl. Hk-1		45,77	15,79			1	.955
Gen. İzl. Hk-2		47,69	16,04				1

Tablo 6. İki Hakemin, İki Farklı Yöntemle Yaptığı Ölçümler Arası Korelasyon Analizi Sonuçları

3.3. Genel İzlenim Yöntemiyle Yapılan Ölçümlerle, DPA Kullanılarak Yapılan Ölçümler Arasındaki Korelasyon

Genel izlenim yöntemi, iki hakem tarafından da uzun bir süre kullanılmış ve halen kullanılmaktadır. İki yöntem ile yapılan ölçümler arasındaki korelasyon, geliştirilen DPA'nın ilgili yapıyı ne düzeyde ölçtüğünü gösterecektir.

Analizde, pearson momentler çarpımı korelasyon katsayısı (r) hesaplanmıştır (Tablo 6). Tablo 6'da görüldüğü üzere, hakem-1'in DPA ile yaptığı ölçümle, genel izlenim yöntemi ile yaptığı ölçüm arasında yüksek korelasyon vardır ($r = .960$). Hakem-2'nin DPA ile yaptığı ölçümle, genel izlenim yöntemi ile yaptığı ölçüm arasında da yüksek korelasyon bulunmuştur ($r = .959$).

SONUÇ, ÖNERİLER VE TARTIŞMA

Araştırma kapsamında geliştirilen "Piyano Eğitimi 1 ve 2 Dersi Performans Ölçümüne Yönelik Dereceli Puanlama Anahtarı" için öncelikle sınıf içi korelasyon katsayısı hesaplanarak hakemler arası tutarlılığa bakılmıştır. Genel puanlar için mükemmel düzeyde, tüm ölçütler içinse iyi düzeyde ve mükemmel düzeyde hakemler arası tutarlılık vardır.

Tüm ölçütler üzerinden hesaplanan Cronbach's Alpha iç tutarlık katsayısı .80'den büyüktür ve "eserde bütünlük ve notasyon" ölçütü hariç, tüm ölçütler DPA'nın güvenilirliğini arttırmaktadır.

"Eserde bütünlük ve notasyon" ölçütü yalnızca bir hakemde güvenilirliği düşürmesinden dolayı bu durum önemli bir anlam ifade etmemektedir.

DPA geliştirilen araştırmalarda, genel güvenilirliği kısmen düşüren ölçütlerle karşılaşmıştır (Burnsed, Hinkle ve King, 1985; Norris ve Borst, 2007) Geliştirilen DPA'nın genel güvenilirliği yüksek düzeydedir. DPA'nın kapsam geçerliği için uzman görüşü alınmıştır. Yapı geçerliği içinse, hakemlerin değerlendirme için kullandıkları genel izlenim yöntemi ile verdikleri puanlarla, DPA kullanılarak verdikleri puanlar arası korelasyona bakılmıştır. Hakem-1 ve hakem-2 için değerlendirme yöntemleri arasında yüksek korelasyon bulunmuştur.

Bu sonuçlar doğrultusunda geliştirilen "Piyano Eğitimi 1 ve 2 Dersi Performans Ölçümüne Yönelik Dereceli Puanlama Anahtarı" kullanımı; (a) sınav anında doldurulabilecek pratik bir ölçme aracı olduğundan, (b) ilgili derslerde, geçerli ve güvenilir bir ölçme aracı kullanılmaması durumunda, sübjektif değerlendirmeler yapılabileceğinden, (c) öğretim elemanının ilgili performansa yönelik beklentilerini açıklamalar yoluyla ifade ettiğinden (d) öğrencilerin ilgili performansa yönelik zayıf ve güçlü yönlerini ortaya çıkararak, öğrencilere yapıcı geri dönütler verilebileceğinden dolayı önerilmektedir. Bu araştırmanın önceliklerinden biri ölçme aracının (DPA) güvenilirliğidir. Cronbach's alpha katsayısı .70 üzeri kabul edilebilir düzey olduğundan (Cortina, 1993), geliştirilen ölçme aracı güvenilirirdir. Geliştirilen DPA, müzik performans ölçümüne yönelik geliştirilen, birçok DPA'dan yüksek düzeyde güvenilirirdir (Azzara, 1993; Barnicle, 1993; Ciorba & Smith, 2009; Levinowitz, 1989; Rutkowski, 1990; Saunders & Holahan, 1997).

Araştırmacılar ülkemizde, Piyano performans ölçümüne yönelik üç araştırma bulmuştur (Kurtuldu, 2010; Grançer Okay 2010; Özalp ve Dalkıran, 2018). Bu araştırmaların ikisi likert tipi ölçek, diğeri analitik türde bir DPA'dır. Araştırmalar incelenmiş, geçerliği ve güvenilirliği sağlanmış olan bu ölçme araçlarının sınav gibi kısıtlı bir zaman içerisinde doldurulabilecek ölçme araçları olmadığı, Holistik DPA türünün, performansın bütününe odaklandığından daha pratik olduğu düşünülmektedir.

Bir ölçüt dışında, madde-bütün analizi sonuçları, ölçütlerin genel güvenilirliği yükselttiği yönündedir. Yalnızca, hakem-2'ye ilişkin analizlerde güvenilirliği düşüren bir ölçüte rastlanmıştır. Ciorba ve Smith (2009) tarafından yapılan araştırmada, tahta üfleli grup-2'de güvenilirliği düşüren iki hakeme; şan grup-8'de ise güvenilirliği düşüren bir hakeme rastlanılmıştır. Burnsed, Hinkle ve King (1985) tarafından yapılan araştırmada, "Diğer (Other)" ölçütü, DPA'nın genel güvenilirliğini kısmen düşürmektedir. Norris ve Borst (2007) tarafından yapılan araştırmada ise, "Ritim (Rhythm)" ölçütü, genel güvenilirliği olumsuz yönde etkilemektedir.

Araştırmada, hakemler arası tutarlılık, iyi düzey (.850) ve mükemmel düzey (.976) arasında değişmektedir. Bergee (1993) araştırmasında, farklı enstrüman grupları için, tutarlılık ve güvenilirlik düzeyleri ile benzer bulgularla karşılaşmıştır. Garman, Boyle, ve DeCarbo (1991) tarafından yapılan araştırmada ise, hakemler arası tutarlılık orta ve iyi düzey arasında değişmektedir.

Genel izlenim yoluyla yapılan ölçümlerle, DPA kullanılarak yapılan ölçümler arasında yüksek düzeyde korelasyon vardır. Dönmez ve Dalkıran (2018), araştırmalarında iki ölçme yöntemi ile alınan sonuçlar arasında yüksek düzeyde uyum bulmuşlardır.

KAYNAKLAR

- Abeles, H. (1973). Development and validation of a clarinet performance adjudication scale. *Journal of Research in Music Education*, 21, 246-255.
- Arı, G. (2008). Öğrencilerin hikâye edici metinlerinin çözümleyici puanlama yönergesine göre değerlendirilmesi (6 ve 7. Sınıf örneği) (Yayımlanmamış Doktora Tezi). Marmara Üniversitesi, Eğitim Bilimleri Enstitüsü, İstanbul.
- Arter, J. A. (2000). *Rubrics, scoring guides and performance criteria: Classroom tools for assessing and improving student learning*. New Orleans: American Educational Research Association.
- Azzara, C. D. (1993). The effect of audiation-based improvisation techniques on musical achievement of elementary music students. *Journal of Research in Music Education*, 41, 328-342.
- Barnicle, S. P. (1993). Evaluating the choral performer. *CMEA News*, 44(2), 24-26.
- Bergee, M. J. (1988). The use of an objectively constructed rating scale for the evaluation of brass juries: A criterion related study. *Missouri Journal of Research in Music Education*, 5(5), 6-25.
- Bergee, M. J. (1989). An investigation of the efficacy of using an objectively constructed rating scale for the evaluation of university-level single reed juries. *Missouri Journal of Research in Music Education*, 26, 74-91.
- Bergee, M. J. (1993). A comparison of faculty, peer, and self-evaluation of applied brass jury performances. *Journal of Research in Music Education*, 41, 19-27.
- Burnsed, V., Hinkle, D., & King, S. (1985). Performance evaluation reliability at selected concert festivals. *Journal of Band Research*, 21(1), 22-29.
- Ciorba, C. R., & Smith, N. Y. (2009). Measurement of instrumental and vocal undergraduate performance juries using a multidimensional assessment rubric. *Journal of Research in Music Education*, 57(1), 5-15.
- Cornett, C. E., & Smithrim, K. L. (2001). *The arts as meaning makers: Integrating literature and the arts throughout the curriculum* (Canadian ed.). Toronto: Pearson Education
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, 78(1), 98.
- Çepni, S. (2011). Performansların Değerlendirilmesi. Emin Karip (Ed.). *Performansların Değerlendirilmesi* (ss.233-285) Ankara: Pegem Akademi.
- Dönmez, L. ve Dalkıran E. (2018). Vurmalı çalgılar dereceli puanlama anahtarı tasarımı ve geleneksel puanlama yöntemi ile ilişki ve farklılığın tespit edilmesi. *Akademik Sosyal Araştırmalar Dergisi*, 84, 500-518.
- Garman, B. R., Boyle, J. D., & DeCarbo, N. J. (1991). Orchestra festival ratings: Interjudge agreement and relationships between performance categories and final ratings. *Research Perspectives in Music Education*, 2, 19- 24.
- Gliem, R. and Gliem, J. (Ekim, 2003), "Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales", in 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education, pp. 82-88.
- Grançer Okay, N. (2010). Piyano performansının değerlendirilmesi (Yayımlanmamış Doktora Tezi). Marmara Üniversitesi, Eğitim Bilimleri Enstitüsü, İstanbul.
- Horowitz, R. A. (1994). The development of a rating scale for jazz guitar improvisation performance. *Dissertation Abstracts International*, 11, 3443.
- Jones, H., Jr. (1986). An application of a facet-factorial approach to scale construction in the development of a rating scale for high school solo vocal performance. *Dissertation Abstracts International*, 47, 1230.
- Kan, A. (2007). Performans değerlendirme sürecine katkıları açısından yeni program anlayışı içerisinde kullanılabilen bir değerlendirme yaklaşımı: Rubrik puanlama yönergeleri. *Kuram ve Uygulamada Eğitim Bilimleri*, 7 (1), 144-152.
- Karasar, N. (2008). *Bilimsel araştırma yöntemi*. (18. Baskı). Ankara: Nobel Akademik Yayıncılık.
- Kırışoğlu, O. T. (2005). *Sanatta eğitim görmek öğrenmek yaratmak*. Ankara: Pegem Akademi Yayınları.
- Kurtuldu, K. (2010). Piyano sınavı değerlendirme ölçeğinin puanlayıcılara yönelik geçerlik ve güvenilirliği. *Elektronik Sosyal Bilimler Dergisi*, 9(31), 224-232.
- Kutlu, Ö., Doğan, C., D. ve Karakaya, İ. (2017). Ölçme ve değerlendirme, performansa ve portfolyoya dayalı durum belirleme. Ankara: Pegem Akademi Yayıncılık.

- Levinowitz, L. M. (1989). *An investigation of preschool children's comparative capability to sing songs with and without words. Bulletin of the Council for Research in Music Education*, 100, 14–19.
- MEB (2007). *Görsel sanatlar dersi (1-8. sınıflar) öğretim programı ve kılavuzu*, Ankara: MEB Devlet Kitapları Müdürlüğü.
- MENC (1996). *Performance standards for music*. Reston: Author.
- Moskal, B. M. (2000). *Scoring rubrics: what, when and how? Practical Assessment in Evaluation*, 7 (3).
- Norris, C. E. & Borst, J. D. (2007). *An examination of the reliabilities of two choral festival adjudication forms. Journal of Research in Music Education*, 55, 237–251.
- Özalp, U. ve Dalkiran, E. (2018). *Piyano performansı dereceli puanlama anahtarının geliştirilmesi. Electronic Turkish Studies*, 13(27).
- Polat, S. (2017). *Ses Eğitiminin Temel Öğeleri ve Çeşitli Söyleme Tekniklerine Yönelik Uzman Görüşleri ve Performans Değerlendirme Ölçeği Örneği. İnönü Üniversitesi Kültür ve Sanat Dergisi*, 3(1), 1-18.
- Portney, L. G., & Watkins, M. P. (2009). *Foundations of clinical research: applications to practice (Vol. 892)*. NJ: Pearson/ Prentice Hall.
- Radocy, R.E. (1986). *On quantifying the uncountable in musical behavior. Bulletin of the Council of Research in Music Education*, 88, 22–31.
- Radocy, R. E. (1989). *Evaluating student achievement. Music Educators Journal*, 76(4), 30–33.
- Rutkowski, J. (1990). *The measurement and evaluation of children's singing voice development. Quarterly Journal of Music Teaching and Learning*, 1(1/2), 81–95.
- Saddler, B. ve Andrade, H. (2004). *The writing rubric. Educational Leadership*, 62 (1), 48-52.
- Saunders, T. C., & Holahan, J. M. (1997). *Criteria specific rating scales in the evaluation of high school instrumental performance. Journal of Research in Music Education*, 45, 259–272.
- Tuna, S. (2011). *Sanat eserlerini değerlendirme. A. O. Alakuş ve L. Mercin (Ed.), Sanat eğitimi ve görsel sanatlar öğretimi (s.253-262)*. Ankara: Pegem Akademi Yayıncılık.
- Uçan, A. (1980). *Çevreden evrene keman eğitimi üzerine. Çağdaş Eğitim Dergisi*, 5(47), 19-27.
- Yarar, B. (2010). *Müzik öğretmenliği lisans programındaki "bireysel ses eğitimi" dersine yönelik performans ölçeği geliştirme çalışması (Yayımlanmamış Yüksek Lisans Tezi)*. Atatürk Üniversitesi, Sosyal Bilimler Enstitüsü, Erzurum.
- Yaşar, M. (2008). *Eğitimde ölçme ve değerlendirmenin önemi. S. Tekinalp (Ed.), Eğitimde ölçme ve değerlendirme (s.1-7)*. Ankara: Pegem Akademi Yayıncılık
- Zdinski, S. F., & Barnes, G. V. (2002). *Development and validation of a string performance rating scale. Journal of Research in Music Education*, 50, 245–255.

PIYANO EĞİTİMİ DERSİ PERFORMANS ÖLÇÜMÜNE YÖNELİK DERECELİ PUANLAMA ANAHTARI

Yönerge: Aşağıdaki dereceli puanlama anahtarı, Piyano Eğitimi dersine yönelik performans ölçümü amacıyla hazırlanmıştır. Öğrencinin performansına göre, “Puan” başlığı altına uygun sayısal ifade yazılır. Tüm ölçütlerden alınan puanlar toplanarak, toplam puan bölümüne yazılır.

Öğrencinin :

Adı Soyadı :

Numarası :

Ölçütler	Yetersiz	Geliştirilmeli	Başarılı	Mükemmel	Puan
	0-4 puan	5-7 puan	8-9 puan	10 puan	
Temel Davranışlar	Performans süresince; nadiren doğru duruş, tutuş, oturuş sürdürüldü ve doğru bilek, kol, vücut hareketleri gerçekleştirildi.	Performans süresince; kısmen doğru duruş, tutuş, oturuş sürdürüldü ve doğru bilek, kol, vücut hareketleri gerçekleştirildi.	Performans süresince; çoğunlukla doğru duruş, tutuş, oturuş sürdürüldü ve doğru bilek, kol, vücut hareketleri gerçekleştirildi.	Performans süresince; doğru duruş, tutuş, oturuş sürdürüldü ve doğru bilek, kol, vücut hareketleri gerçekleştirildi.	___/10
Dizi, Kadans ve Arpej (DKA)	DKA; nadiren bütünlük içerisinde, istenilen hızda, doğru parmak numaraları ve geçişleri ile seslendirildi.	DKA; kısmen bütünlük içerisinde, istenilen hızda, doğru parmak numaraları ve geçişleri ile seslendirildi.	DKA; çoğunlukla bütünlük içerisinde, istenilen hızda, doğru parmak numaraları ve geçişleri ile seslendirildi.	DKA; bütünlük içerisinde, istenilen hızda, doğru parmak numaraları ve geçişleri ile seslendirildi.	___/10
Etüt(ler)					
Etüt(ler)te Bütünlük ve Notasyona Uygunluk	0-9 puan Etüt(ler); nadiren bütünlüğe ve notasyona uygun şekilde seslendirildi.	10-15 puan Etüt(ler); kısmen bütünlüğe ve notasyona uygun şekilde seslendirildi.	16-19 puan Etüt(ler); çoğunlukla bütünlüğe ve notasyona uygun şekilde seslendirildi.	20 puan Etüt(ler); bütünlüğe ve notasyona uygun şekilde seslendirildi.	___/20
Etüt(ler)te Çalm Teknikleri	0-4 puan Etüt(ler); nadiren çalm tekniklerine uygun şekilde seslendirildi.	5-7 puan Etüt(ler); kısmen çalm tekniklerine uygun şekilde seslendirildi.	8-9 puan Etüt(ler); çoğunlukla çalm tekniklerine uygun şekilde seslendirildi.	10 puan Etüt(ler); çalm tekniklerine uygun şekilde seslendirildi.	___/10
Etüt(ler)te Müzikalite	0-4 puan Etüt(ler); nadiren müzikal becerilere uygun şekilde seslendirildi.	5-7 puan Etüt(ler); kısmen müzikal becerilere uygun şekilde seslendirildi.	8-9 puan Etüt(ler); çoğunlukla müzikal becerilere uygun şekilde seslendirildi.	10 puan Etüt(ler); müzikal becerilere uygun şekilde seslendirildi.	___/10
Eser(ler) / Küçük Ölçekli Parça(lar)					
Eser(ler)de Bütünlük ve Notasyona Uygunluk	0-9 puan Eser(ler); nadiren bütünlüğe, notasyona ve dönem özelliklerine uygun şekilde seslendirildi.	10-15 puan Eser(ler); kısmen bütünlüğe, notasyona ve dönem özelliklerine uygun şekilde seslendirildi.	16-19 puan Eser(ler); çoğunlukla bütünlüğe, notasyona ve dönem özelliklerine uygun şekilde seslendirildi.	20 puan Eser(ler); bütünlüğe, notasyona ve dönem özelliklerine uygun şekilde seslendirildi.	___/20
Eser(ler)de Çalm Teknikleri	0-4 puan Eser(ler); nadiren çalm tekniklerine uygun şekilde seslendirildi.	5-7 puan Eser(ler); kısmen çalm tekniklerine uygun şekilde seslendirildi.	8-9 puan Eser(ler); çoğunlukla çalm tekniklerine uygun şekilde seslendirildi.	10 puan Eser(ler); çalm tekniklerine uygun şekilde seslendirildi.	___/10
Eser(ler)de Müzikalite	0-4 puan Eser(ler); nadiren müzikal becerilere uygun şekilde seslendirildi.	5-7 puan Eser(ler); kısmen müzikal becerilere uygun şekilde seslendirildi.	8-9 puan Eser(ler); çoğunlukla müzikal becerilere uygun şekilde seslendirildi.	10 puan Eser(ler); müzikal becerilere uygun şekilde seslendirildi.	___/10
Toplam Puan					___/100