Elementary Education Online, 2018; 17(3): pp. 1463-1475
İlköğretim Online, 2018; 17(3): s.1463-1475. [Online]: http://ilkogretim-online.org.tr
doi 10.17051/ilkonline.2018.466366

Ortaokul Öğrencilerine Yönelik Blok Flüt İcra Performansı Dereceli Puanlama Anahtarının Güvenirliğinin Genellenebilirlik Kuramı İle İncelenmesi

The Study on the Reliability of the Grading Key Measuring the Performance of the Block Flute Performance of the Secondary School Students Via Generalizability Theory

Alican Gülle, *Mersin Cengiz Topel Ortaokulu*, *alicangulle@gmail.com*
Nezaket Bilge Uzun, *Mersin Üniversitesi Eğitim Fakültesi*, *n.bilgeuzun@gmail.com*
Cenk Akay, *Mersin Üniversitesi Eğitim Fakültesi*, *cenkakay35@hotmail.com*

Öz. Araştırmanın amacı, genellenebilirlik kuramı ile farklı değişkenlik kaynaklarından da gelen hatalar dikkate alınarak ortaokul öğrencilerinin blok flüt icra performanslarını ölçen, geçerli ve güvenilir bir ölçme aracı geliştirmektir. Yapılan araştırmada öğrencilerin blok flüt icra performansına yönelik performans düzeylerini ölçen ve analitik dereceli puanlama anahtarından elde edilen performans puanlarının güvenirliğinin belirlenmesi amaçlandığından betimsel bir araştırmadır. Araştırmanın çalışma grubunu 6. Sınıfta öğrenim gören 23 öğrenci oluşturmaktadır. Öğrencilerin blok flüt icra performansı araştırmacılar tarafından hazırlanan analitik dereceli puanlama anahtarı ile üç müzik öğretmeni tarafından puanlanmıştır. Analitik dereceli puanlama anahtarı ile toplanan verilerin analizinde, tüm hata kaynaklarını aynı anda değerlendirerek güvenirliğin belirlenmesini sağlayan Genellenebilirlik Kuramı (GK) kullanılmıştır. GK ya dayalı yapılan analizler sonucunda; ortaokul öğrencilerine yönelik oluşturulan blok flüt icra performansı dereceli puanlama anahtarının güvenilir ve geçerli (G: 0.89, Phi: 0,79) bir ölçme aracı ve ideal puanlayıcı sayısının 4 olduğu; kapsam geçerliği bakımından belirlenen görev basamaklarının beceriyi yeterince açığa çıkardığı sonucuna ulaşılmıştır.

Anahtar Sözcükler: Blok flüt performansı, Genellenebilirlik teorisi, Müzik eğitimi, Analitik rubrik

Abstract. In this regard, the main purpose of the research is to develop a valid and reliable measurement tool which measures the performance of the block flute performance of the secondary school students, considering the errors from generalizability theory and different sources of variability. It is a descriptive study as it is aimed to measure the performance levels of the students for block flute performance and to determine the reliability of their performance scores obtained from the analytical grading key. The study group consists of 23 students in the 6th grade. The block flute performance of the students wasscored by three music teachers with the analytical grading key prepared by there searchers. Generalizability Theory (GT) which helps determine the reliability by evaluating all error sources at the same time was employed in the analysis of the data collected with the analytical grading key. It was found as a result of analyzes based on GT that the analytical grading key for the block flute performance of secondary school students is reliableandvalid (g: 0.89, phi: 0.79), the ideal scorernumber is 4, and the task steps determined in terms of content validity are able to discover the skill.

Keywords: Block flute performance, Generalizability theory, Music instruction, Analytic rubric

Geliş tarihi: 11.01.2018 Kabul tarihi: 14.05.2018 Yayımlanma tarihi: 31.07.2018

SUMMARY

Introduction

It is of great importance that the students' musical performance as a result of learning processes regarding the music lesson is measured and evaluated as accurately as possible. When the MNE Music Teaching Program is examined, the basic learning areas "listening - speaking - playing" require the use of a different performance scale for each learning area. Therefore, developing an analytical grading key in the measurement for block flute performance is thought to divide the performance into items, score it more precisely and reveal the learning skills of the students clearly and reliably. The main purpose of the research is to develop a valid, reliable and analytical grading key which measures the performance of the block flute performance of the secondary school students.

Method

During the preparation of the grading key for block flute performance, area specialists were collaborated to ensure the construct and content validity. For these basic skills, sub-tasks and convenience statements corresponding to these basic skills are specified for both content and construct validity together with area specialists. 5 sub-tasks (note, note period, metronome, technique, integrity) for the ability to play, 5 sub-tasks (note, note period, metronome, technique, integrity) "deciphering" and 6 sub-tasks (metronome, note period, note, harmony, attention, exercise) were determined and convenience statements corresponding to these sub-task were written in a 4-point Likert form. To conduct comprehensive reliability analyzes and, by interpreting the sources of variability, to obtain findings as to the reliability and validity of the grading key, Generalizability Theory (GK), which involves strong statistical techniques such as classical test theory and variance analysis, and enables to determine the reliability by evaluating all error sources at the same time, was employed. The content validity rates were determined by asking a measurement evaluation specialist and four music field specialists to score the data collection tool. It was made ready for implementation by making final arrangements in line with the validity rates and experts' recommendations.

Results

According to the results of the generalizability study affecting the block flute performance that was revealed by the grading key of which content validity was ensured based on expert opinions, it is concluded that the highest variance of all predicted variance percentages is the variance percentage (b) for the individual. Therefore, the variability in the measurement results related to performance is the students who are the main object of the measurement, which refers to the reliability and validity of the measurement tool. The second highest variance component in the main effects belongs to the scorer (14%), indicating that the scorers are different in their generosity and stiffness in scoring. The lowest variance percent predicted in the main effects belongs to the items (m). This suggests that there is no differentiation between items in terms of difficulty.

Regarding the variance components of interaction effects, the highest variance component is the variance component belonging to the individual and item interaction (bm) while the scorer and the individual interaction (bp) is the second highest interaction effect. This result shows that scorers may be a systematic error source because they act biased from one individual to the other. The fact that the variance component predicted for the scorer and item interaction effect (pm) is relatively low indicates that the scorers consistently scored the items.

Examining the predicted residual variance component (bpm, e), it is supposed that there are random errors in the evaluation of performance and that different sources of variability (scorer, gender, student's ability, interest, etc.) which do not exist in the design have effect on the result. In this study, the fact that percentage of the predicted residual variance component is lower than of the basic object of measurement can be interpreted that the most significant variable that may affect the performance exist in the design, and there are possible sources of random errors, all of which prove the design applied in the generalization study is determined correctly.

When G (0.89) and Phi (0.79) coefficients obtained as a result of generalizability study are examined, it is seen that the criteria, skills and tasks in the measuring tool are reliable and valid.

The results of the decision studies (K studies) through the variance components obtained from generalizability studies show that evaluation of block flute performance by four scorers would make G and phi coefficients the most ideal and also bring more reliable results. On the other hand, decision-making results for item numbers point that the criteria, convenience statements and qualities of the grading key consisting of 16 task numbers are specified well enough.

Discussion and Conclusion

Findings of the study show that the developed measurement tool gives reliable and generalizable results in determining the block flute performance of the students. The finding found in generalizability studies that the highest estimated variance component belongs to the students who are the main object of the measurement is one of the main indicators for revealing individual differences in performance by the measurement tool, which proves the measurement tool is able to well-distinguish individual performances. The fact that the items in the grading key have the lowest percentage of variance indicates that the items and the criteria associated with the items are well defined. This finding also provides evidence for the construct and content validity of the measurement tool used for evaluating performance. Considering the scorers can both differ in the level of stiffness-generosity during scoring and give biased points, it is expected that G and phi coefficients will increase in case of decreasing the systematic errors caused by the scorer and increasing the number of scorers. The obtained G and Phi coefficients are within the acceptance boundaries, indicating that the grading key used for performance evaluation can also be used for similar situations. In this regard, it is considered that the use of the performance scale developed for the block flute performance evaluations in schools can provide a realistic measurement, identify the weaknesses and strengths of the performances of the students, and reveal the success of the education process at the same time.

GİRİŞ

Müzik duygusal ifade gücünü gösteren sesleri sözlü veya sözsüz birleştirme çabasıdır. Aristo'nun duyguları ifade etme konusunda hiçbir şeyin ritim ve şarkı söylemek kadar güçlü olmadığını düşünmesi ve müziğin küçük yaştan itibaren eğitimin tüm kademelerindekullanılması

gerektiğini savunması, müziğin eğitimde önemli bir yere sahip olduğunu göstermektedir (Türkmen 2016). Ülkemizde müzik eğitiminin verildiği ilk kurum ilköğretimdir. İlköğretim içerisinde yer alan ortaokul, ilkokulun devamı olan ve öğrencilerin alan uzmanlarıyla (branş öğretmenleri) duyuşsal, bilişsel, psikomotor öğrenme alanları dikkate alınarak öğretim sürecinin devamını sağlayan eğitim kurumudur. Özellikle öğrencilerin müzik dersine karşı geliştirdikleri duyuşsal (tutum, ilgi, değer) değerler ilk defa alan uzmanı eşliğinde oluştuğundan ve bu durumun öğrencilerin yaşam boyu müzik değerlerine etki edebileceği düşünüldüğünden ortaokulda verilen müzik eğitimi büyük önem taşımaktadır.

MEB 2017 eğitim programları raporu incelendiğinde müzik dersi öğretim programının temel yapısı 1.sınıftan 8.sınıfa kadar ortak bir temel yapı üzerine kurulmuş, disiplinler arası etkileşimin açık olduğu bir yaklaşım çerçevesinde; "Dinleme - Söyleme - Çalma", "Müziksel Algı ve Bilgilenme", "Müziksel Yaratıcılık", "Müzik Kültürü " olarak dört temel öğrenme alanı üzerine oluşturulmuştur (MEB 2017). "Dinleme-Söyleme-Çalma" temel öğrenme alanı içerisinde yer alan "Çalma" temel öğrenme alanı, müzik öğretmeninin gerçekleştirmiş olduğu "çalgı eğitimi" sonucunda öğrencinin bilişsel-duyuşsal ve devinişsel öğrenmelerini bir araya getirmesi ve müziksel bir performans ortaya koyması anlamına gelmektedir. "Çalma", ilköğretim müzik öğretim programında blok flüt enstrumanı ile gerçekleştirilmektedir. Blok flütün; kullanım, taşıma ve maddi avantaj bakımından sağladığı kolaylıklar nedeniyle müzik öğretim programında tercih edildiği düşünülmektedir. Blok flüt eğitiminde bilişsel-duyuşsal ve devinişsel öğrenmeler bir arada gerçekleştiğinden, enstruman eğitimine başlamadan önce öğrencilere temel müzik eğitiminin verilmesi gerekmektedir. Temel müzik eğitimi ile öğrencilere temel müzik kavramları (porte, anahtar, oktav, sol anahtarı, solfej, deşifre, bona, nota, ritim vb.) öğretilir ve öğrenciler enstruman eğitimine hazır hale getirilir.

Öğretici notaların blok flütteki yerlerini gösterir ve öğrencilerle birlikte üfleyerek çalışmayı gerçekleştirir. İnce sesten kalın sese doğru ilerledikçe nefes şiddetinin önemi vurgulanarak, notaların doğru bir şekilde üflenmesi sağlanır. Ardından öğretilecek olan şarkı-marş-türkünün deşifre çalışması gerçekleştirilir. "Deşifre", hiç bilinmeyen bir eserin notalarının okunması veya çalınmasıdır. Öğrenciler yapmış oldukları deşifre çalışması ile temel müzik eğitiminde almış oldukları öğrenmeleri tekrarlar ve müziksel öğrenmelerinin kalıcı olmasını sağlarlar. Deşifre çalışması sırasında öğrencinin; nota, nota süresi, eserin ritim hızı vb. yanlışları varsa tespit edilir, düzeltilir ve eserin deşifresinin blok flütle yapılması sağlanır. Her bir ölçü tek tek ele alınarak deşifre çalışması gerçekleştirilir. Ardından sınıf gruplara ayrılır ve her grup ölçüleri birleştirerek eseri bir bütün halinde çalar. Çalışma tekrarlanır ve her tekrar sırasında grup içerisindeki öğrencilerin hataları düzeltilir. Tüm gruplarla çalışma gerçekleştirildikten sonra öğretici yönetiminde sınıf, "toplu uygulama icra"sı gerçekleştirerek eseri bir bütün halinde çalıp, müziksel performanslarını sergiler.

Öğrencilerin gerçekleştirdiği müziksel performans, yaşamış oldukları öğrenme süreçlerinin gözlemlenmesini ve başarı durumunu ortaya koymaktadır. Bu nedenle, müziksel performansın müzik öğretmeni tarafından doğru, dikkatli bir şekilde ölçülmesi ve değerlendirilmesi büyük önem taşımaktadır. Bu tip performanslara dayalı yapılan ölçme değerlendirme işlemlerinin güvenirliği ve geçerliği açısından kullanılan ölçme araçları önem taşımaktadır. Performans değerlendirilmesinde açık uçlu sorular, gözlem formları, kontrol listeleri, dereceleme ölçekleri, dereceli puanlama anahtarları (rubrikler) vb. ölçme araçları kullanılabilmektedir. Performans değerlendirme, öğrencilerin bilgi ve becerilerini kullanarak cevap vermelerini sağlayan, geçerliği ve güvenirliği yüksek ölçme araçlarıyla yapılan değerlendirme biçimidir (Büyüköztürk, 2007). Performans değerlendirmede öğrenciden cevabı oluşturması ya da bilgisini bir ürün ortaya koyarak göstermesi istenir (Stiggins, 1994).

Müziksel performansın ölçülüp değerlendirilmesinin tarihsel süreçte çok eskilere dayanmadığı, bu durumun yirminci yüzyılın ortalarından itibaren müzik performans ölçeklerinin uyarlanmasına, uygulanmasına ve metot geliştirilmesine neden olduğu görülmektedir (Saraç, Şeker 2009). Geliştirilen metotlarda performans ölçümü, dinleyenlerin doğal tepkileri olarak hoşlanma/ hoşlanmama, dinlemeye ayrılan süre ve duygusal olarak kesinlikle hoşlanma/ kesinlikle hoşlanmama şeklinde gerçekleştirilmektedir (Le Blank, 1998: 425. akt, Saraç, Şeker 2009). Öğrencilerin performanslarına yönelik değerlendirme çalışmaları özneldir ve çok çeşitli

faktörler tarafından etkilenmektedir (Woods, 1997). Bu açıklamalar doğrultusunda müziksel performansının değerlendirmesinde hangi tür performans ölçeğinin kullanılması gerektiğini seçmek, önem taşıyan bir durumdur. Dereceli puanlama anahtarı (Rubrik), performans değerlendirmelerinde yaygın olarak kullanılan ölçme araçlarındandır. Popham (1997; akt, Parlak ve Doğan 2014) dereceli puanlama anahtarını her bir çalışma için ölçütleri listeleyen ve çalışmada nelerin yer aldığını belirten bir ölçme aracı olduğunu ifade etmiş; değerlendirme ölçütleri, ölçüt tanımlamaları ve bir puanlama stratejisi olmak üzere 3 bölümden oluştuğunu belirtmiştir.

1. Değerlendirme ölçütleri: Kabul edilebilir yanıtları, kabul edilemez yanıtlardan ayırmak için kullanılır.

2. Ölçüt tanımlamaları: Öğrencilerin yanıtlarındaki niteliksel farklılıkları tanımlama biçimini ifade eder.

3. Puanlama stratejisi: Puanlamada, analitik (analitical) ya da bütünsel (holistic) dereceli puanlama anahtarları kullanılabilir. Analitik puanlama anahtarında performansı oluşturan unsurlar, tanımlanmış ölçütler doğrultusunda birbirinden bağımsız olarak puanlanıp kaydedilir (Haladyna, 1997; Moskal, 2000). Bütünsel puanlama anahtarında ise performansın farklı düzeylerinin ortaya çıkarılması için belirlenmiş ölçütler arasında bir ayrışma bulunmadığından öğrencinin gösterdiği performans bütün olarak belirlenip tek puan verilmektedir. (Brookhart, 1999; Kutlu, Doğan, Karakaya, 2009).

Araştırma konusu ile ilgili yapılan alan yazın taraması sonucunda; blok flüt icra performansına yönelik bir ölçek çalışmasının yapılmadığı, farklı enstruman performanslarına yönelik ölçek çalışmalarının yapıldığı ve genel olarak enstruman icra performanslarına yönelik ölçek çalışmalarının yeterli olmadığı gözlenmiştir.

Dalkıran (2008) "Keman eğitiminde performansın ölçülmesi" başlıklı çalışmasında, müzik eğitimi ana bilim dallarında verilen keman eğitimi için yarıyıl sonu sınavlarında kullanılabilecek bir ölçme aracı geliştirmiştir. Araştırmanın örneklemini 6 devlet üniversitesinin eğitim fakültelerine bağlı güzel sanatlar eğitimi bölümlerinde dersi yürüten 23 öğretim elemanı ile bu dersi alan 330 keman öğrencisi oluşturmaktadır. Ölçme aracı "Program Boyutu, Sınav Performans Boyutu, Yarıyıl İçi Durum Boyutu" olmak üzere 3 ana boyut; "programın gereklerine uygunluk, doğru ve temiz ses üretimi, keman çalmaya hazır bulunma, sağ ve sol el tekniği, metrik ve ritmik doğruluk, artikülasyon, bütünlük, ton kalitesi, hız ve gürlük, vibrato, dönem içi performans" olmak üzere11 alt boyuttan oluşturulmuştur. Yapılan istatistiksel analiz sonucunda geliştirilen ölçeğin iç tutarlılık katsayısı0.87 olarak belirtilmiştir.

Kurtuldu ve Çiftçi (2010) "Yaylı çalgılar performans değerlendirme ölçeği geçerlik ve güvenirlik analizi" başlıklı çalışmasında yaylı çalgılar dersine yönelik performans değerlendirme ölçeği geliştirmiştir. Araştırmacılar tarafından oluşturulan değerlendirme basamakları göz önünde bulundurularak, yaylı çalgılar için temel olan 8 basamak dikkate alınıp ölçek oluşturulmuştur. Araştırmanın çalışma grubunu 2 devlet üniversitesinin müzik eğitimi ana bilim dalında öğrenim gören 163 öğrenci oluşturup geliştirilen ölçek, 6 öğretim elemanı tarafından yılsonu sınavlarında öğrencilerin değerlendirilmesinde kullanılmıştır. Yapılan istatistiksel analiz sonucunda ölçeğin iç tutarlılık katsayısı 0,96 olarak belirtilmiştir.

Saraç ve Şeker (2010) "Güzel sanatlar eğitimi bölümlerinde çalgı öğretimindeki performansın değerlendirilmesi" başlıklı çalışmasında Güzel Sanatlar Eğitimi Bölümleri Müzik Eğitimi Anabilim Dalı programlarındaki yaylı çalgı eğitiminin daha nesnel ve bilimsel veriler ile değerlendirilmesi için geliştirilmeye açık bir ölçek modeli tasarlamıştır. Araştırmanın örneklemini Kazım Karabekir Eğitim Fakültesi Müzik Eğitimi Ana Bilim Dalında yaylı çalgı öğrenimi gören 76 öğrenci ile uygulamaya katılan 3 öğretim elemanı oluşturmaktadır. 26 sorudan oluşan ölçme aracı 5'li likert tipinde hazırlanmış ve uygulanmıştır. Elde edilen veriler sonucunda öğretim elemanlarının farklı değerlendirmeler içerisinde oldukları, puanlayıcılar arasındaki farklılaşmanın azaltılabilmesi için, yaylı çalgı performans değerlendirme ölçeklerinin geliştirilmesi ve standardizasyona ihtiyaç duyulduğu belirtilmiştir.

Akçay (2011) "Gitar eğitiminde performans ölçeği geliştirme çalışması" başlıklı yüksek lisans çalışmasında bir ölçme aracı oluşturmuş, 2 devlet üniversitesinin müzik eğitim ana bilim dalında öğrenim görmekte olan 7 gitar öğrencisi üzerinde uygulamıştır. 4 öğretim elemanının puanlayıcı olarak katıldığı araştırma sonucunda geliştirilen ölçme aracının Cronbach's Alfa

güvenilirlik katsayısı 0.84, Kendall's W puanlayıcılar arası güvenilirlik katsayısı 0.60 olarak belirtilmiştir

Yukarıda sunulan ölçekler incelendiğinde; geliştirilen performans ölçeklerinin birbirinden farklı enstruman performanslarını değerlendirdiği görülmektedir. Bu bağlamda da MEB Müzik Öğretim Programında yer alan "Dinleme - Söyleme - Çalma" temel öğrenme alanı, her öğrenme alanı için farklı bir performans ölçeği kullanımını gerektirmektedir. "Çalma" temel öğrenme alanı içerisinde verilen blok flüt eğitimi sonucunda öğrenci, derste öğrenmiş olduğu bir eseri (çalma), yeni bir eseri (deşifre) ve toplu uygulama icrasıyla bir eseri icra edebilir. Dolayısıyla; öğrencinin blok flüt çalma performansını belirleyebilmek için, üç beceriyi karşılayan ve puanlamasını sağlayan bir ölçme aracı kullanımı gerekmektedir. Bu nedenle blok flüt icrasına yönelik performans ölçümünde dereceli puanlama anahtarı geliştirilmesinin; performansın ögelere ayrıştırılıp daha ayrıntılı puanlanmasını ve öğrencilerin öğrenme becerilerini belirgin, güvenilir bir şekilde ortaya koymasını sağlayacağı düşünülmektedir.

Performans değerlendirmede kullanılacak ölçme araçlarının geliştirilme çalışmaları genellikle puanlayıcılar arası güvenirliğe dayalı yapılmaktadır. Puanlayıcı, performans değerlendirmenin güvenirliğini düşüren önemli bir hata kaynağı olmakla birlikte; puanlayıcının; görev, zaman, cinsiyet, öğrenme- öğretim sürecindeki değişiklikler vb. gibi farklı faktörlerle olan etkileşiminin de en az o kadar önemli bir hata kaynağı olabileceğini düşündürmektedir. Bu nedenle performansa dayalı ölçme araçlarının güvenirlik çalışmalarında hata kaynakları arasındaki etkileşimin de göz önünde bulundurulması gerekmektedir. Performans değerlendirmenin güvenirliği, diğer hata kaynaklarıyla olan etkileşimi birlikte ele alan ve ölçmenin üç temel kuramından biri olan, genellenebilirlik kuramı (GK)' na dayalı yöntemlerle çalışılabilmektedir (Güler, 2008; Güler ve Taşdelen, 2015). Bu nedenle bu çalışmada; ölçme aracının güvenirlik ve geçerlik bulguları ile güvenirlik ve geçerlik arasındaki farkı ortadan kaldıran, ölçme aracının genellenebilirliğine vurgu yapan GK' ya dayalı analizler gerçekleştirilmiştir. Geliştirilen bu ölçme aracının temellerinin GK ile araştırılması, olası farklı değişkenlik kaynakları ve etkileşimlerinin de dikkate alınması nedeniyle daha güvenilir ve geçerli sonuçlar ortaya koymaktadır. Aynı zamanda genellenebilirlik çalışması sonuçlarından yola çıkılarak analizlerin ikinci aşaması olan karar çalışmaları ile; belirlenen ölçütlerin yeterliliği, puanlayıcı sayısı… vb. hakkında araştırmacılara benzer çalışmalar için öneri sunan bir yapı sergilemektedir.

Bu bağlamda araştırmanın temel amacı, genellenebilirlik kuramı ile farklı değişkenlik kaynaklarından da gelen hatalar dikkate alınarak ortaokul öğrencilerinin blok flüt icra performanslarını ölçen, geçerli ve güvenilir bir ölçme aracı önermektir. Bu temel amaca bağlı olarak; performansa dayalı bir ölçümün güvenirliğini düşürebilecek olası hata kaynaklarına dikkat çekmek ve öğrencilerin blok flüt çalma performansını değerlendirmede en güvenilir değişkenlik koşullarını belirlemek amaçlanmıştır. Yapılan literatür taraması sonucunda blok flüt icra performansına yönelik bir ölçme aracının bulunmaması ve rubrik geliştirmede genellikle puanlayıcılar arası güvenirliğe odaklanılması sebebiyle araştırmanın önemli ve özgün olduğu düşünülmektedir.

Araştırmanın Problemi ve Alt problemleri

"Ortaokul öğrencilerinin blok flüt icra performansına yönelik performans düzeylerini ölçen analitik dereceli puanlama anahtarının güvenirliğinin genellenebilirlik kuramı ile incelenmesi; bazı yüzeylerin koşul sayılarının değişimlenerek karar çalışmasının yapılması" amacı doğrultusunda aşağıda yer alan sorulara cevap aranmıştır.

1. Ortaokul öğrencilerinin blok flüt icra performansı analitik dereceli puanlama anahtarı için birey, puanlayıcı, madde ve bunların etkileşimlerine ait kestirilen varyans bileşenleri nasıldır?
2. Ortaokul öğrencilerinin blok flüt icra performansı analitik dereceli puanlama anahtarı değerlendirmesi sonucunda elde edilen puanlamaların, güvenirlik (G ve Phi) katsayıları nasıldır?
3. Puanlayıcı ve madde sayısının değişimlenmesinin G ve Phi katsayılarına olan etkisi nasıldır?

YÖNTEM

Araştırmanın Türü

Bu araştırma; ortaokul öğrencilerinin blok flüt icra performans ölçümlerine ilişkin farklı değişkenlik kaynakları işe koşularak kapsamlı güvenirlik analizlerinin yürütüldüğü; blok flüt icra performansı dereceli puanlama anahtarının geliştirilmesi, ideal ölçme durumları için uygun koşul sayılarının belirlenmesi ve uygulamalı araştırmalar için temel oluşturabilecek bir ölçme aracının alana kazandırılmasını amaçlaması açısından temel araştırma niteliğindedir.

Çalışma Grubu

Araştırmanın çalışma grubunu 6. Sınıfta öğrenim gören 23 öğrenci oluşturmaktadır. Öğrencilerin blok flüt icra performansı araştırmacılar tarafından hazırlanan analitik dereceli puanlama anahtarı ile üç müzik öğretmeni tarafından bağımsız bir biçimde puanlanmıştır.

Veri Toplama Aracı

Blok flüt icra performansı dereceli puanlama anahtarında yer alan performans adımlarının kapsamının belirlenmesi için MEB müzik öğretim programında yer alan kazanımlar ve literatürde yer alan farklı enstrumanların icrasına ilişkin gözlemlenen adımlar incelenmiştir (Akçay 2011; Dalkıran 2008; Kurtuldu ve Çiftçi 2010;MEB 2017;Saraç ve Şeker 2010). Müzik dersi programında önerildiği üzere blok flüt icrası performansının değerlendirilmesinde "çalma, deşifre, toplu icra" becerilerinin ölçülmesi gerektiğine karar verilmiştir. Analitik rubriğin hazırlanması aşamasında kapsam ve yapı geçerliğini sağlayabilmek için alan uzmanları ile birlikte çalışılmıştır. Bu temel beceriler için hem kapsam geçerliği hem de yapı geçerliği için iki müzik alan uzmanı ile becerilere karşılık gelen alt görevler ve uygunluk ifadeleri belirlenmiştir. "Çalma" becerisine yönelik 5 (nota, nota süreleri, metronom, teknik, bütünlük), "deşifre" beceresine yönelik 5 (dikkat, nota süreleri, nota, teknik, egzersiz), "toplu icra" becerisine yönelik 6 (metronom, nota süreleri, nota, uyum, dikkat, egzersiz) olmak üzere toplam 16 alt görev belirlenmiş; görevlere karşılık gelen uygunluk ifadeleri 4'lü likert biçiminde düzenlenmiştir. Oluşturulan dereceli puanlama anahtarının anlaşılırlığı, uygulanacak hedef kitleye uygunluğu, yazılan ölçütlerin ve ölçütler için belirlenen görev ifadelerinin uygunluğu vb. için; uygulama öncesi bir ölçme değerlendirme uzmanı ve dört müzik öğretmenine dereceli puanlama anahtarında yer alan her bir hücre puanlatılmıştır. Dereceli puanlama anahtarı için elde edilen uzman görüşleri arasındaki uyum/uyumsuzluk düzeylerinden yola çıkılarak oluşturulan ölçme aracının kapsam geçerlik oranları ve kapsam geçerlik indeksi Lawshe tekniğinin uzantısı olan düzeltilmiş kappa uyum istatistiklerinin hesaplanması yoluyla değerlendirilmiştir. Böylelikle dereceli puanlama anahtarının kapsam geçerliğine ilişkin nicel kanıtlar elde edilmiştir. Hücreler için hesaplanan düzeltilmiş kappa uyum istatistikleri (görev bazında kapsam geçerlik oranları) 0,52 ile 1,00 arasında değişmekte olup, tüm dereceli puanlama anahtarı için kapsam geçerlik indeksi ise 0,95 olarak belirlenmiştir. Son olarak; görev bazında elde edilen kapsam geçerlik oranları dikkate alınarak görev 3, görev 4 ve görev 18 için belirtilen öneriler doğrultusunda son düzenlemeler yapılarak geçerliği sağlanan dereceli puanlama anahtarı uygulamaya hazır hale getirilmiştir.

Verilerin Toplanması

Çalışma grubunu oluşturan öğrenciler; 2 farklı ezgi için performans sergilemiştir. 1. Ezgi öğrencilerin müzik dersinde öğrendikleri bir ezgi iken, 2. ezgi öğrencilerin öğrenmelerine uygun olarak müzik öğretmeni tarafından bestelenen 8 ölçülük bir ezgidir. Aynı zamanda 1.ezgi ile 5 erli gruplar halinde öğrencilerin toplu icra gerçekleştirmesi sağlanmıştır. Bu sayede öğrenci performansları, blok flüt çalma performansında yer alan 3 beceriye (çalma, deşifre, toplu icra) yönelik hazırlanan dereceli puanlama anahtarı doğrultusunda, 3 puanlayıcı tarafından bağımsız bir şekilde değerlendirilip puanlanarak veriler toplanmıştır.

Verilerin Analizi

Bu çalışmada; temelinde klasik test kuramı ve varyans analizi gibi güçlü istatistiksel teknikleri barındıran ve tüm hata kaynaklarını aynı anda değerlendirerek güvenirliğin

belirlenmesini sağlayan Genellenebilirlik Kuramı (GK) kullanılmıştır. GK ölçme sonuçlarının güvenirliğinin belirlenmesini, güvenilir gözlemlerin tasarımını, araştırılmasını ve kavramsallaştırılmasını sağlayan; Cronbach ve arkadaşları tarafından geliştirilen istatistiksel bir kuramdır (Brennan, 2001; Cronbach, Gleser, Nanda ve Rajaratman, 1972). Alan yazındaki ölçme aracı geliştirme çalışmalarının çoğunun temelleri klasik test kuramına bağlı olmakla birlikte, bu kurama bağlı olan yöntemler güvenirliğin anlamına ve ele alınan hata kaynağına göre farklılıklar içermektedir (Lord ve Novic, 1968). Maddelerin temel hata kaynağı olduğu, madde ve toplam test puanı arasındaki ilişkilere dayalı hesaplanan güvenirlik katsayısı; iç tutarlılık, puanlayıcıların hata kaynağı olup; puanlamalar arasındaki ilişkilere dayalı hesaplanan güvenirlik; tutarlılık ve objektiflik, zamanın temel hata kaynağı olduğu ve bir testin farklı zamanlarda uygulanma sonucu elde edilen puanlar arasındaki ilişkiyi inceleyen güvenirlik katsayısı ise kararlılık anlamlarına gelebilmektedir (Crocker ve Algina, 1986; Shavelson ve Webb, 1991; Brennan, 2001). Oysa tek bir analizde farklı anlamları barındıran bu güvenirlik değerleri elde edilemeyeceği gibi, farklı güvenirlik yöntemleri ile hesaplanan güvenirlik katsayılarının değerleri de farklılaşmaktadır. Bu durumda hangi güvenirlik katsayısının kullanılmasının doğru olabileceği tartışılabilir. GK ise aynı anda maddelerden, puanlayıcılardan, farklı zamanlardaki ölçümlerden vb. oluşabilecek olası hata kaynaklarını tek bir analizde ele alabilmekte, bu hata kaynaklarından gelebilecek tesadüfi ve sistematik hatalar hakkında araştırmacılara yorum yapabilme olanağı sunmaktadır.

GK'da, güvenirliğin araştırılmasında Genellenebilirlik çalışması (G-çalışması) ve Karar çalışması (K-çalışması) olmak üzere iki aşama söz konusudur (Goodwin, 2001). G-çalışmasında, üzerinde çalışılan örneklemin evrene genellenebilmesi için, çalışmada yer alan olası tüm değişkenlik kaynakları (varyans bileşenleri) ve etkileşimleri ANOVA yöntemi kullanılarak kestirilir. Kestirilen bu varyans bileşenleri bir sonraki aşama olan K-çalışmasında optimize edilerek en uygun değişkenlik kaynaklarının koşulları belirlenmeye çalışılır. Bir diğer ifadeyle K-çalışması "Eğer ki bu olursa ne olur?" sorusuna cevap aramak üzere yapılan bir çalışmadır (Alharby, 2006). Bu çalışmada yapılan karar çalışmalarında, puanlayıcı ve dereceli puanlama anahtarında beceriler için tanımlanmış görev sayıları değişimlenerek elde edilen G ve Phi katsayılarıincelenmiş; belirlenen görev sayılarının ve kullanılan puanlayıcı sayısının elde edilen güvenirlik değerlerine ne kadar katkı sunduğu araştırılmıştır.

Bu çalışmada; her bir öğrencinin (b) her bir görevi (m) gerçekleştirdiği ve her bir puanlayıcının (p) da her bir öğrencinin her görevini puanladığı; tümüyle çaprazlanmış "bxpxm" desen kullanılmıştır. Veri analizi için EDU-G programı kullanılmıştır.

BULGULAR ve TARTIŞMA

Bu bölümde araştırmanın alt problemlerine yönelik üç alt başlık halinde bulgular sunulmuş ve tartışmaları yapılmıştır.

Ortaokul öğrencilerinin blok flüt icra performansı analitik dereceli puanlama anahtarı için birey, puanlayıcı, madde ve bunların etkileşimlerine ait kestirilen varyans bileşenleri nasıldır?

3 puanlayıcının 23 öğrenciyi blok flüt icra performansı dereceli puanlama anahtarında yer alan 16 görev doğrultusunda puanlamasıyla oluşturulmuş bxpxm deseni, yedi varyans kaynağına ayrılmaktadır. Bu varyans kaynaklarına ait kestirilen genellenebilirlik kuramı ile elde edilen varyans bileşenleri Tablo 1 de açıklanmıştır.

Tablo 1. *G çalışması sonucu elde edilen varyans bileşenleri*

Varyans Kaynağı	Kareler Toplamı	Df	Kareler Ortalaması	%
B	462.35326	22	21.01606	33.9
P	125.13225	2	62.56612	14.0
M	63.99909	15	4.26661	2.9
Bp	80.07609	44	1.81991	8.3
Bm	223.64674	330	0.67772	11.4
Pm	46.95471	30	1.56516	4.8
Bpm	187.83696	660	0.28460	24.7
Toplam	1189.99909	1103		100%

Genellenebilirlik analizi ile elde edilen varyans bileşenlerine ait bulgular ana etkiler, etkileşim etkileri ve artık etki biçiminde yorumlanmıştır.

Ana etkiler dikkate alındığında; bireylere ilişkin varyans yüzdesinin en yüksek değere sahip olduğu görülmektedir. Öğrenciler için kestirilen varyans (b) bileşeni ölçmeye konu olan özellik açısından bireyler arası farklılıklardır. Bu araştırmada öğrenciler ölçmenin temel objesidir. Genellenebilirlik kuramına bağlı olarak yürütülen çalışmalarda ölçmenin temel objesinden kaynaklanan değişkenliğin fazla olması beklenmektedir (Güler, 2008). Çünkü performansa ilişkin ölçme sonuçlarındaki değişkenliğin temel nedeninin bireyler, yani ölçmenin temel konusu olması beklenir. Öğrenciler için kestirilen varyans (b) bileşenin en yüksek değere sahip olması (%33,9), bireylerin söz konusu performans bakımından birbirinden oldukça farklılaştığının, heterojenleştiğinin göstergesidir. Bir diğer ifadeyle, performansa dayalı kullanılan analitik puanlama anahtarı ile yapılan ölçümlerde bireysel farklılıklar ortaya çıkarılmıştır. Bu söz konusu ölçme aracının güvenirliğine ve geçerliğine dair bir kanıt olarak düşünülebilir.

Ana etkiler içinde en yüksek ikinci bileşen puanlayıcılara aittir (%14). Puanlayıcı değişkenlik kaynağına ilişkin varyans (p),puanlayıcıların objektifliği, tutarlılığı hakkında bilgi vermektedir. Ancak puanlayıcılara ilişkin varyans yüzdesinin yüksek olması, puanlayıcılardan kaynaklı bir sistematik hatanın varlığına işaret etmektedir. Puanlayıcı ana etkisine bağlı sistematik hata genellenebilirlik çalışmalarında, puanlayıcıların puanlamadaki katılık ve cömertlikleri biçiminde yorumlanmaktadır.

Tüm ana etkilerde en düşük varyans yüzdesi görevlere(m) aittir (%2,9). Bu durum beceriyi oluşturan görevlerde güçlük ve zorluk bakımından bir farklılaşmanın bulunmadığını göstermektedir. Aynı zamanda görevlerin en düşük varyans yüzdesine sahip olması, görevlerin ve görevlerle ilişkilendirilen ölçütlerin oldukça iyi tanımlandığının bir göstergesidir. Nitekim yapılan karar çalışmaları da bu durumu desteklemektedir (bkz. Tablo 2). Bu bağlamda söz konusu dereceli puanlama anahtarı için belirlenen görevlerin performansı ölçmede güvenilir ve geçerli sonuçlar üreteceği düşünülmektedir.

Etkileşim etkilerine ait varyans yüzdeleri dikkate alındığında en yüksek varyans bileşeninin öğrenci ve görev etkileşimine ait olan varyans bileşeni (bm) olduğu görülmektedir (11,4). Öğrenciler performans ölçütlerini oluşturan maddelerde sergiledikleri performanslar açısından farklılıklar sergilemektedir. Öğrenci ve görev etkileşiminin yüksek çıkmasının; öğrencilerin icra ettiği ezgiye aşina olması-olmaması, önceden icra etmesi-etmemesi, icra edilen ezgiyi sevmesi-sevmemesi, ezginin müzik türünü sevmesi-sevmemesi, blok flüt enstrumanına ilgi duyması-duymaması…vb. karıştırıcı değişkenlerinden kaynaklanabileceği düşünülmektedir. Yanısıra bu etkileşim etkisine ait varyans bileşeninin yüksek çıkmasının bir diğer nedeni; bu etkileşimi oluşturan değişkenlerden birinin birey olması ve ana etkiler içinde en yüksek varyans bileşeninin bireylere ait olması olabilir.

Puanlayıcı ve öğrenci etkileşimi (bp) dikkate alındığında, bu etkileşimin en yüksek 2. etkileşim etkisi olduğu görülmektedir (%8,3). Bu sonuç puanlayıcılarla öğrenciler arasında bir etkileşim olduğu ve puanlayıcıların öğrencileri puanlarken sistematik hata yaptığını göstermektedir. Bir diğer ifade ile bir puanlayıcı öğrenciden öğrenciye farklı değerlendirmeler yapmış, yanlı davranmıştır. Puanlayıcıların öğrencileri puanlamada sistematik hatayı en az seviyeye düşürmesinin, etkileşim etkisi sonucunu değiştireceği ve daha hatasız sonuçlar ortaya

koyabileceği düşünülmektedir. Aynı zamanda bu yüzdenin yüksek çıkmasın, ana etkilerdeki puanlayıcılara ait yüzdenin yüksek olmasından kaynaklanabileceği söylenebilir.

Puanlayıcı ve görev (pm) etkileşim etkisinin diğer değişkenlik kaynaklarına oranla görece düşük olması, puanlayıcıların maddeleri tutarlı puanladığı biçiminde yorumlanabilir. Görevlerden kaynaklanan ana etkinin düşük olmasının, etkileşim etkisinin düşük çıkmasına neden olduğu söylenebilir. Elde edilen sonuca göre; puanlayıcıların görevler veya ölçütlerle önceden bir yaşantı geçirmediği, söz konusu performansa bir bütün olarak baktığı, uzmanlık alanlarına göre maddelerle etkileşim içerisine girmediği ve puanlamada hata yapmadığı düşünülmektedir.

Artık varyans sonucu incelendiğinde yapılan ölçmede tesadüfi hataların var olduğu ve desende bulunmayan farklı değişkenlik kaynaklarının da (puanlayıcı cinsiyeti, birey becerisi, ilgi, yaşantı) sonuca etki ettiği düşünülmektedir. Bir çok desende; artık etki, tesadüfi hata ve farklı değişkenlik kaynaklarından gelen yüzdenin en yüksek değere sahip olduğu görülmektedir. Bu bağlamda kestirilen artık varyans bileşeninin ölçmenin temel objesi olan öğrencilerden daha düşük çıkması, bu performansı etkileyebilecek en önemli değişkenlik kaynaklarının desen içerisinde yer aldığını ve olası tesadüfi hata kaynaklarının var olduğu biçiminde yorumlanabilir.

Ortaokul öğrencilerinin blok flüt icra performansı analitik dereceli puanlama anahtarı değerlendirmesi sonucunda elde edilen puanlamaların, güvenirlik (G ve Phi) katsayıları nasıldır?

G (0.89) ve Phi (0.79) katsayılarına bakıldığında elde edilen sonuç ölçme aracında yer alan ölçütlerin, becerilerin ve görevlerin güvenilir ve geçerli olduğunu göstermektedir. Öğrencilere bağlı olan ayırt edicilik varyansının yüksek olmasının, en yüksek varyans yüzdesinin ölçmenin temel objesine ait olmasının ölçme aracının geçerliği için kanıt olarak kullanılabileceği; puanlayıcıların daha tutarlı puanlama yapmasının G ve Phi katsayılarını arttıracağı düşünülmektedir.

Puanlayıcı ve madde sayısının değişimlenmesinin G ve Phi katsayılarına olan etkisi nasıldır?

G çalışması kapsamında incelenen varyans bileşenleri yorumlanmış ve incelenmiş, bu doğrultuda karar çalışmaları puanlayıcı ve madde sayısı için yürütülmüştür. Puanlayıcı sayısı 2, 4, 5, 6, 7 biçiminde; görev sayısı da K çalışmaları kapsamında 10, 12, 14, 18, 20 biçiminde değişimlenmiş ve her bir değişimlenme için G ve Phi katsayıları elde edilmiştir. Tablo 2 de karar çalışmaları sonuçları yer almaktadır.

Tablo 2. *Madde ve puanlayıcı karar çalışması*

Madde Sayısı (m)		**Görev Sayısı (m)**					
		16*	**10**	**12**	**14**	**18**	**20**
	G Katsayısı	**0.89**	0.88	0.89	0.89	0.90	0.90
	Phi Katsayısı	**0.79**	0.78	0.78	0.79	0.79	0.80
Puanlayıcı Sayısı (p)		**Puanlayıcı Sayısı (m)**					
		3*	**2**	**4**	**5**	**6**	**7**
	G Katsayısı	0.89	0.86	**0.91**	0.93	0.94	0.94
	Phi Katsayısı	0.79	0.72	**0.83**	0.86	0.87	0.89

*G çalışmasında kullanılan değişkenlik kaynaklarının koşul sayıları

Tablo 2 deki K çalışmaları sonuçları incelendiğinde; puanlayıcı sayılarında artış veya düşüş yapıldığında G ve Phi katsayılarında değişmeler oluştuğu, 4 puanlayıcı ile gerçekleştirilecek ölçmenin G ve Phi katsayılarını en ideal boyuta getireceği ve daha güvenilir sonuçlar ortaya koyacağı düşünülmektedir. Görev sayıları üzerinde yapılan karar çalışması sonuçları incelendiğinde,16 görevden oluşan dereceli puanlama anahtarında yer alan ölçütlerin, uygunluk ifadelerinin ve niteliklerin iyi belirlendiği sonucuna ulaşılmıştır.K çalışmalarında söz konusu performansın değerlendirilmesinde puanlayıcı sayısının değişimlenmesinin, G ve Phi katsayılarına sunmakta olduğu katkı fazla bulunmuştur. Bu sonuçta; özellikle puanlayıcıların kullanıldığı performans değerlendirmelerinde puanlayıcının önemli bir hata kaynağı olduğunu, puanlayıcı güvenirliklerine bağlı olarak geliştirilen ölçme araçlarının hatalı sonuçlar

doğurabileceğini düşündürmektedir. Öte yandan görev sayılarında artış veya düşüşe ihtiyaç duyulmadığı; bu bağlamda dereceli puanlama anahtarında belirlenen beceriler ve becerilere ait alt görevlerin kapsam geçerliğinin sağlandığı düşünülebilir.

SONUÇ

Yapılan araştırma sonucunda elde edilen bulgular, geliştirilen ölçme aracının, öğrencilerin blok flüt icra performansını belirlemede güvenilir ve genellenebilir sonuçlar verdiğini göstermektedir. Bu bağlamda okullarda blok flüt icra performansı değerlendirmelerinde geliştirilen performans ölçeğinin kullanılmasının; gerçekçi bir ölçme yapmayı sağlayacağı, öğrencilerin performanslarındaki zayıf ve güçlü yönlerini belirleyerek öğrencinin almış olduğu enstruman eğitim sürecinin başarı düzeyini göstereceği düşünülmektedir.

- Ortaokul öğrencilerine yönelik oluşturulan blok flüt icra performansı dereceli puanlama anahtarı güvenilir ve geçerli bir ölçme aracıdır (G: 0.89, Phi: 0,79).
- Ortaokul öğrencilerine yönelik oluşturulan blok flüt icra performansının değerlendirilmesinde önerilecek ideal puanlayıcı sayısı 4'tür.
- Ortaokul öğrencilerine yönelik oluşturulan blok flüt icra performansı dereceli puanlama anahtarında yer alan beceri ve bu becerilerin altında yer alan görevler uygun bir biçimde belirlenmiştir. Diğer bir ifadeyle; kapsam geçerliği bakımında belirlenen görev basamakları beceriyi yeterince açığa çıkarmaktadır.
- 3 temel alt görev açısından oluşturulan dereceli puanlama anahtarı performansa ilişkin farklılıkları yeterince ayırtedebilmektedir.

ÖNERİLER

Uygulamaya Yönelik Öneriler

- Ortaokul öğrencilerine yönelik oluşturulan blok flüt icra performansı dereceli puanlama anahtarı, Milli Eğitim Bakanlığı tarafından Müzik Öğretim programında ölçme aracı olarak yer verilip müzik öğretmenleri tarafından kullanımı yaygınlaştırılabilir.
- Blok flüt icra performansı dereceli puanlama anahtarında yer alan üç nitelik (çalma, deşifre, toplu uygulama), müzik öğretmenlerinin ihtiyacına göre öğrencilerin performans düzeylerini belirlemede ayrı ayrı kullanılabilir.
- Geliştirilen performans ölçeği, Melodika gibi üflemeli enstrumanlara değişikliklerle uyarlanabilir.

İleride Araştırma Yapacaklar İçin Öneriler

- Ortaokul öğrencilerinin blok flüt icrasına yönelik tutumlarını ölçen bir tutum ölçeği araştırmacılar tarafından geliştirilebilir.
- Ortaokul öğrencilerinin blok flüt icrasına yönelik öz yeterliklerini ölçen bir öz yeterlik ölçeği araştırmacılar tarafından geliştirilebilir.
- Geliştirilen bu ölçme aracı kullanılarak blok flüt çalma performansını etkileyen ve bu performans üzerinde etkili olabileceği düşünülen başka değişkenler ile çalışılabilir.
- Öğrencilerin cinsiyeti, puanlayıcının cinsiyeti, okul türü, müzik öğretmeninin nitelikleri… vb. gibi farklı değişkenlik kaynakları ile genellenebilirlik çalışmaları ve karar çalışmaları yürütülebilir.
- Farklı ezgiler işe koşularak benzer çalışmalar yapılabilir.

- İleride yapılabilecek benzer araştırmalara kaynak oluşturma ya da yardımcı olma bağlamında kullanılabilir.

KAYNAKÇA

Akçay, Ş.Ö.(2011). *Gitar eğitiminde performans ölçeği geliştirme çalışması.* Atatürk Üniversitesi Eğitim Bilimler Enstitüsü Güzel Sanatlar Eğitimi Anabilim Dalı Müzik Öğretmenliği Bilim Dalı Yüksek lisans tezi

Alharby, E.R. (2006). A comparison between two scoring methods, holistic vs. Analytic using two measurement models, the Generalizability Theory and the many facet Rasch measure men twith in the context of performance assessment. *Unpublished doctor aldissertation. The Pennsylvania State University Faculty of Education, Pennsylvania*

Atılgan, H. , Kan, A. ve Doğan, N. (2006). *Eğitimde Ölçme ve Değerlendirme.* Ankara: Anı Yayıncılık.

Baykul, Yaşar. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması.* Ankara: ÖSYM Yayınları, Cem Web Ofset, 89.

Brennan, R. L. (2001). *Generalizabilitytheory*. Iowa City, IA: ACT Publications.

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*

Cronbach, J. L., Gleser, G. C., Nanda, H., & Rajaratman, N. (1972). *The dependability of behavioral measurements: Theory of generalizability for scores and profiles.* New York: John Wiley and Sons.

Çiftçi, E. Kurtuldu, K.M. (2010). Yaylı çalgılar performans değerlendirme ölçeği geçerlik ve güvenirlik analizi. *Erzincan Eğitim Fakültesi Dergisi Cilt-Sayı: 12-2.*

Dalkıran, E. (2008). Keman eğitiminde performansın ölçülmesi. *Yüzüncü Yıl Üniversitesi, Eğitim Fakültesi Dergisi Cilt:V, Sayı:II, 116-136*

Goodwin, L. D. (2001). Interrater agreement and reliability. *Measurement in Psychical Education and Exercises Science, 5(1), 13-14.*

Güler, N., Uyanık, K. G. ve Teker, T. G. (2012). *Genellenebilirlik kuramı.* PegemA Yayıncılık. Ankara

Güler N., Taşdelen Teker G. (2015). Açık Uçlu Maddelerde Farklı Yaklaşımlarla Elde Edilen

Puanlayıcılar Arası Güvenirliğin Değerlendirilmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, Cilt 6, Sayı 1, 12-24.*

Haladyna, M. T. (1997). *Writing test items to evaluate high erorder thinking.* Need ham Heights: Allyn and Bacon

Kadıoğlu H, ADA Sefer. (2009). *"Türk Eğitim Sistemi"*

Kutlu, Ö., Doğan, D. C. ve Karakaya, İ. (2009). *Öğrenci başarısının belirlenmesi: performansa ve portfolyaya dayalı durum belirleme.* Ankara: Pegem Akademi.

Le Blank, A.,Jin,Y.C. vd. (1998), "Pictorial versus Verbal Rating Scales in Music Preference Measurement" *Journal of Research in Music Education v46 no3 p425-35*

MEB (2017) *"İlköğretim Müzik Dersi Öğretim Programı"*

Moskal, B., M. (2000*).* Scoringrubrics: what, when, how? *PraticalAssessment, Research and Evaluation, 8 (14).*

Popham, J. W. (1997). What's wrong and what'sright with rubric. *Educational Leadership. 55, (2), 12*

Parlak, B. Doğan, N. (2014). Dereceli Puanlama Anahtarı ve Puanlama Anahtarından Elde Edilen Puanların Uyum Düzeyleri. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi. 29(2), 189-197*

Saraç, G.,Şeker, H. (2010). Güzel sanatlar eğitimi bölümlerinde çalgı öğretimindeki performansın değerlendirilmesi. *Dergipark , Sayı 20 cilt: 0*

Shavelson, R. J.,&Webb, N. M. (1991). *Generalizabilitytheory: A primer.* California: Sage Publications.

Stiggins, R. J. (1994). *Student-Centered classroom assessment*. New York: Macmillan Publishing Company.

Türkmen, E.F. *Müzik eğitiminde öğretim yöntemleri.* Ankara: Pegem akademi yayıncılık.