

İlkokul Düzeyinde Yapılan Öğretmen, Öz ve Akran Değerlendirmelerine İlişkin Güvenirliğin Genellebilirlik Kuramı ile İncelenmesi *

Eda GÜRLER **

Nagihan BOZTUNÇ ÖZTÜRK ***

Emel EMİNOĞLU ****

Öz

Bu çalışmada ilkokul düzeyinde yapılan öğretmen, öz ve akran değerlendirmelerine ilişkin güvenilirlik katsayılarının belirlenmesi amaçlanmıştır. Bu amaç doğrultusunda disiplinler arası yaklaşım temel alınarak değerler eğitimi kapsamındaki yardımseverlik konusu Türkçe, Hayat Bilgisi ve Müzik derslerine yönelik uygulamalarla ilişkilendirilerek ele alınmıştır. Araştırmanın çalışma grubunu; Ankara ilinde bulunan bir devlet okulunun üçüncü sınıfında okuyan 30 öğrenci oluşturmaktadır. Çalışmanın amacı doğrultusunda verilerin güvenilirlik analizinde Genellebilirlik Kuramı kullanılmıştır. Çalışmanın sonucunda; her üç derste de öğrenci ana etkisi için kestirilen varyans bileşeninin toplam varyansın en büyük bileşeni olmuştur. G ve Φ katsayılarına bakıldığında; müzik .80'in üzerinde; Türkçe ve Hayat Bilgisi derslerinde ise .90'ın da üzerinde güvenilirlik katsayıları elde edilmiştir. G-Yüzeyleri analizi sonuçlarına göre ise; öğretmen ve akran değerlendirmelerinin sırasıyla analizden çıkarılmasıyla G ve Φ katsayıların düşme eğilimine, öz değerlendirmenin analizden çıkarılmasıyla ise bu katsayılar yükselme eğilimine sahip olmuştur. Özellikle Müzik dersinde öğretmen ve akran değerlendirmelerinin analizden çıkarılmasıyla elde edilen güvenilirlik katsayılarının .60 civarına gerilemesi oldukça dikkat çekici bir sonuçtur.

Anahtar Kelimeler: Öğretmen değerlendirmesi, öz değerlendirme, akran değerlendirmesi, genellebilirlik kuramı

GİRİŞ

Eğitim sisteminin önemli bir ögesi olan değerlendirme; öğretim sürecinin etkililik ve verimlilik düzeyi hakkında bilgi verme, belirlenen hedeflerin kazandırılma derecesini saptama, uygulamaların güçlü ve zayıf yönlerini ortaya koyma gibi önemli işlevlere sahiptir. Değerlendirme etkinlikleri doğru uygulandığında, eğitimin sürekli kontrolünü sağlayarak, her aşamada ortaya çıkan aksaklıkların hızlı bir şekilde düzeltilmesini ve problemlere sağlıklı çözüm yolları üretilmesini mümkün kılar. Öğrenci gelişiminin takip edilmesiyle, öğrenme güçlüklerinin ve eksikliklerinin belirlenip giderilebilmesine imkân tanır. Başarı ve başarısızlık kaynaklarının tespitiyle eğitimi olumlu ve olumsuz etkileyen unsurların açığa çıkarılmasına yardımcı olur. Böylece hem eğitim kalitesini artıran uygulamaların desteklenmesi hem de engel ve tehditlere karşı zamanında önlem alınması mümkün olur. Gelecek için yapılacak planlama ve yönlendirme çalışmalarına da ışık tutularak eğitimin verimli ve hızlı bir şekilde gelişmesi sağlanır (Çeçen, 2011; İşman & Eskicumalı, 2003; Kurudayıoğlu, Şahin & Çelik, 2008; Turgut & Baykul, 2015; Yaşar, 2017).

Gerçekleştirilen tüm eğitim etkinliklerinin yarattığı değişimi her yönüyle ortaya koyabilen bir değerlendirme yapabilmek için öğretmenler farklı yöntemler kullanmaktadırlar. Bu yöntemler sonucunda öğrenciler, öğretmenlerin veya uzmanların hazırladığı ölçütler doğrultusunda, onların bakış açılarıyla değerlendirilmektedirler. Ancak eğitim ve öğretim etkileşimle gerçekleşen süreçlerdir. Bu etkileşime aktif olarak katılan öğrencilerin bakış açısının değerlendirme etkinliklerinde yer almaması

* Bu makalenin bir kısmı 17. Sınıf Öğretmenliği Eğitimi Sempozyumunda sözlü bildiri olarak sunulmuştur.

** Doç. Dr., Hacettepe Üniversitesi, Eğitim Fakültesi, Ankara-Türkiye, edaerdem@hacettepe.edu.tr, ORCID ID: 0000-0002-1719-9840

*** Öğr. Gör. Dr., Hacettepe Üniversitesi, Yaşam Boyu Öğrenme Merkezi, Ankara-Türkiye, nagihanboztunc@hacettepe.edu.tr, ORCID ID: 0000-0002-2777-5311

**** Öğretmen, Talatpaşa Ortaokulu, Ankara-Türkiye, emel.eminoglu@gmail.com, ORCID ID: 0000-0003-1712-8652

Bu makaleye atıfta bulunmak için:

Gürler, E., Boztunç-Öztürk, N., & Eminoğlu, E. (2019). İlkokul düzeyinde yapılan öğretmen, öz ve akran değerlendirmelerine ilişkin güvenilirliğin genellebilirlik kuramı ile incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 10(4), 406-421. doi: 10.21031/epod.583891

Geliş Tarihi: 28.06.2019

Kabul Tarihi: 21.08.2019

önemli bir eksiklik meydana getirir. Bu şekilde yapılan değerlendirme etkinlikleri, sürece katılmayan öğrenciler için yeterince anlamlı hale gelmeyecek dolayısıyla da işlevini tam olarak gerçekleştiremeyecektir. Değerlendirme etkinliklerinin anlamlı hale getirilmesi, öğrencilerin değerlendirme kriterlerini hem kendi çalışmalarına hem de başka çalışmalara uygulamaları ile mümkündür. Bu şekilde öğrenciler değerlendirme işleminin derin bir öğrenme tecrübesi olduğunun farkına varabilirler. Bir etkinlikle ilgili çalışmalarını diğer öğrencilerin aynı etkinlikle ilgili çalışmaları ile karşılaştırarak hem daha derinlemesine öğrenebilirler hem de değerlendirme sürecinde zihnin çalışma prensiplerini anlayabilirler. Böylece öğretmenin değerlendirme işlemini nasıl yaptığı konusunda da fikir sahibi olabilirler. Bu durum öğretmen öğrenci diyalogunun yolunu açabileceği gibi öğrencinin değerlendirme sonrasında yapılacak düzenlemeler üzerinde düşünmesini ve sorumluluk almasını da sağlar. Kendi öğrenme süreçleriyle ilgili sorumluluk alan öğrenciler; düşünen, yönlendiren, kendi gelişiminin farkına varabilen, kendi çalışmasını organize edebilen, kendini eleştirebilen, bağımsız öğrenenler olma fırsatını yakalarlar. Sergilediği bir davranışın, o davranışla ilgili ölçütlere uygun olup olmadığı konusunda karar verme becerisine sahip olan bir birey, yaşamı boyunca herhangi bir otoriteden bağımsız olarak kendi davranışlarını kontrol etme yetisine de sahip olacaktır. Dolayısıyla değerlendirme etkinlikleri, yaşam boyu öğrenme özerkliği kazanmış bireyler yetiştirilmesine katkı sağlayacaktır (Race, 2001; Sünbül, 2007; Wilson & Jan, 1993).

Değerlendirme etkinliklerinin bu yönünü keşfeden eğitim dünyası belirgin bir şekilde bireyin öğrenmesinin bireyin merkezde olduğu yöntemlerle değerlendirilmesine yönelmektedir. Bu tür değerlendirmeler daha maliyetli ve zaman alıcı olmalarına rağmen öğrenmeyle değerlendirmeyi bütünleştirerek hem öğrencilerin öğrenmelerine hem de öğretmenlerin mesleki gelişimine önemli oranda katkıda bulunmaktadır. Bu etkinlikler yalnızca bir puanlama çalışmasını değil aktif katılımı öğrenme becerilerinin geliştiği ve derinlemesine öğrenmenin gerçekleştiği dinamik bir süreci ifade etmektedir. Öğrencilerin bu sürece katılmaları, değerlendirmenin yalnızca bir not verme işlemi olmadığını anlamalarını sağlamaktadır. Değerlendirmenin amaç ve işlevleri konusunda yeterince bilgilendirilmeyen öğrenciler, öğretim etkinliklerinin ulaşmak istediği noktaları tam olarak anlayamayabilirler. Kendisinden beklenenler hakkında yeterince bilgi sahibi olmayan bir öğrencinin öğrenme motivasyonu olumsuz yönde etkilenebilmektedir. Bu durum öğrenmeye karşı yanlış tutumlar geliştirilmesine yol açabilmektedir. Etkinliğin amacını ve gerekliliğini anlayan öğrenciler, değerlendirme kriterlerine daha gerçekçi yaklaşarak, güçlü ve zayıf taraflarını keşfedebilmektedirler. Kendini keşfedebilen öğrenciler kendi öğrenmelerinin sorumluluğunu alarak doğrudan öğrenmeye odaklandıklarında; özgüveni yüksek, eleştirel ve bağımsız öğrenenler olma yolunda ilerlemektedirler (Alıcı, 2010; Ballantyne, Huges & Mylonas, 2002; Boud, 1986; Cihanoğlu, 2008; Cram, 1995; Falchikov, 2001; Tekindal, 2014; Topping, Smith, Swanson & Elliot, 2000).

Bu tür değerlendirme etkinliklerinin nesnel bir şekilde gerçekleşerek amacına ulaşabilmesi için öğrencilerin sürece değerlendirmenin ilk basamaklarından itibaren dahil olmaları gerekir. Öğrenciler, değerlendirme türünün tespit edilmesi, hangi öğrenme çıktılarının değerlendirileceğinin belirlenmesi ve kullanılacak ölçütlerin oluşturulması aşamalarına aktif olarak katılmalıdırlar. Bu konularda verilecek kararları öğretmen ve öğrenciler birlikte tartışıp uzlaşarak vermelidirler. Değerlendirme konusunda sürece dahil olanlar arasında bir uyum sağlanmalıdır. Böylece öğrenciler kendilerinden beklenen ideal davranışların, bu davranışların beklenme nedenlerinin ve öğrenmenin gerekliliğinin farkına varabilirler. Bir davranışa yönelik ölçüt oluşturabilme ve davranışın kalitesini derecelendirme becerileri gelişir. Öğrencilerin bu tartışmalarda yer alması, iletişim ve kendini ifade etme becerileri açısından da faydalı olacaktır. Bütün bu öğrenilenlerle öğrenciler kendi öğrenme süreçlerini baştan sona yönetebilirler. Öğrenme seviyelerini yükseltmek için yapılması gerekenlere kendileri karar verebilirler (Alıcı, 2010; Stiggins & Chappius, 2005; Woolfolk 2002).

Öğrencilerin değerlendirme etkinliklerine katılması, sağlıklı bir öğretim-öğrenme ortamı yaratılmasına da katkı sağlamaktadır. Bu etkinlikler, öğretmene öğrencinin nasıl düşündüğünü, dolayısıyla da nasıl öğrenebileceği konusunda bilgi verirler. Öğretmenin öğrencileri duyuşsal özellikleri de içeren farklı yönleriyle tanınmasını sağlarlar. Böylece öğretmene öğretim etkinliklerinin düzenlenmesinde rehberlik ederler. Aynı zamanda öğrencinin de öğretmenin nasıl düşündüğünü anlamasına yardımcı olurlar. Öğrenciler benzer düşünme yollarını kullanarak sürece dahil

olduklarında öğrenme ortamının bir parçası haline geldiklerini hissederler. Potansiyellerini fark eden öğrencilerin akademik benlik kavramı olumlu yönde gelişir; özgüvenleri artar. Başarılarını ve zaman içindeki gelişim düzeylerini görerek öğrendikleri ile gurur duyarlar. Böylece, mutlu ve öğrenmeye karşı istekli olurlar (Bahar, 2006; Stiggins & Chappius, 2005).

Öğrenci katılımı ile gerçekleşen deđerlendirme etkinlikleri, öğrencilerin kendi öğrenmelerinin farkına vardıktan sonra diđer öğrencilerin öğrenmelerinin de farkına varmasına olanak sağlamaktadır. Öğrenciler kendilerinden başlayarak birbirlerinin de öğrenme sorumluluđunu almakta ve diđer bireyleri deđerlendirme becerisi kazanmaktadırlar. Öğrencilerin bu gelişimini sağlayan iki önemli deđerlendirme türü öz deđerlendirme ve akran deđerlendirme olarak karşımıza çıkmaktadır.

Öz deđerlendirme, öğrencilerin deđerlendirme kriterlerini kendi çalışmalarına uygulayarak bu kriterleri ne derece yerine getirdikleri hakkında yargıda bulunmalarıdır. Böylece öğrenciler ne bildiklerini ne yapabildiklerini ve nasıl öğrendiklerini keşfederler. Bu keşif sürecinde eleştirel bir bakış açısıyla üst düzey düşünme becerilerini kullanma fırsatı yakalayan öğrencilerin, kendilerini nesnel bir biçimde anlamlandırmaları sağlanır. Öğrenciler güçlü ve zayıf yanlarını tanıyarak yaşadıkları öğrenme sorunlarının farkına varırlar. Kendi öğrenme yolları hakkında edindikleri detaylı bilgileri kullanarak kendi öğrenme sorunlarına çözüm üretebilirler. Kazandıkları öğrenme yaşantılarını yargılayarak gelecekteki işlerini ve çalışmalarını planlama yetilerini geliştirirler. Dolayısıyla kendi başarısını deđerlendirme becerisi kazanmış bir öğrencinin, daha büyük başarıları elde edecek yetkinliğe ulaşacağı söylenebilir. Bu nedenle öğrencilerin öz deđerlendirme sonucunda ortaya çıkan dönütlere dayanarak bir hareket planı ile üretici ve gerçekçi amaçlar seti oluşturmalarına yardım edilmelidir (Alıcı, 2010; Boud, 1986; Kutlu, Dođan & Karakaya, 2008; Mistar, 2011; Stiggins, 1997; Tekindal, 2014).

Bilişsel ve yapılandırmacı öğrenme kuramları açısından bakıldığında, öz deđerlendirmenin bilginin öğrenen tarafından yapılandırılmasına yardımcı olduđu görülmektedir. Bu kuramlara göre, yeni bilgilerin öğrenci için anlamlı hale gelmesi var olan eski bilgilerle ilişkilendirilmesine bağlıdır. Öz deđerlendirme, öğrencilere özüksedikleri ölçütlere dayanan anlamlı geri bildirimler vererek var olan bilgi ve anlayış ile yenisi arasında bağ kurulmasına katkı sağlamaktadır. Bu şekilde öğrenciler kendi bilgi ve anlayışlarını öğrenme hedefleri ile sürekli karşılaştırarak öğrenirler. Bu durum öz deđerlendirmenin öğrenme hedefi yönelimi oluşturmada da etkili olduğunu göstermektedir. Öğrenme hedefleri belirli bir ölçüde bilginin içsel olarak işlenmesini gerektirir. Öz deđerlendirme, içsel kontrol, bilgi, anlayış ve becerileri geliştirerek öğrencilerin bilgiyi tam olarak anlamaya yönelik gelişimlerinin farkında olmalarını sağladığı için öğrenme türü motivasyonlarına katkı yapmaktadır (McMillan, çev. 2015).

Öz deđerlendirme bireyin yansıtma yetisinin gelişimi ile yakından ilişkilidir. Yansıtma, kişinin dışarıdan bir gözlemci gibi kendini izlemesini ve gelecekte daha iyi eylemlerde bulunmak için karar verme becerilerini geliştirmesini içerir (Osterman & Kottkamp, 1993). Öğrencilerin yansıtıcı davranış ve becerilerindeki gelişmeler öz deđerlendirmedeki en önemli noktadır. Bu konuda gelişme sağlanabilmesi için hangi davranış ve becerilerin deđerlendirileceđi ile bunlara karşılık gelen eğilimlere ilişkin açıklamaların anlaşılır bir şekilde tanımlanması gerekir. Öğrencilerin çalışmaları ile ilgili yansıtıcı yorumlar edinebilmek için onlardan beklenenler açık bir şekilde ifade edilmelidir. Bu konuda eğilimleri görselleştiren basit örnekler kullanılabilir. İşe öğrencilerin dersle ilgili sorulara verdikleri cevapların doğruluđu sorgulanarak başlanabilir. Devamında ise cevabın niçin doğru olmadığı, yanlış cevabın öğrenciye tam olarak ne anlattığı, doğru cevap verebilmek için yapılması gerekenlerin neler olduđu gibi sorular yöneltilebilir (McMillan, çev. 2015).

Öz deđerlendirme araçları farklı biçimlerde hazırlanabilir. Öğrencinin neleri öğrendiđini ve hangi alanlarda sorun yaşadığını belirleyerek, öğrenmelerinin sorumluluđunu almalarını sağlamak koşuluyla; taslak biçiminde hazırlanmış kontrol listelerinden ve sorulardan başlayan bir formattan, bir kompozisyon yazısında ürettikleri yansımaları sorgulayan formata kadar deđişiklik gösterebilir (Bahar, Nartgün, Durmuş & Bıçak, 2008). Öğrencilerin yapmış oldukları öz deđerlendirmeler, öğrencilerin bireysel gelişim dosyalarında saklanır (Woolfolk, 2002).

Öz değerlendirmenin sağlıklı bir biçimde yapılmasını engelleyen birtakım faktörler de bulunmaktadır. Nesnel yorumlama yapmanın zorluğunu yaşayan öğrencilerin kendi öğrenmelerini değerlendirmede yanlı davranması; kendi yeteneklerini olduğundan fazla görmesi veya küçümsemesi; kendi yeteneklerinin farkında olmadığı için öz değerlendirme yapamaması; kendisini, öz değerlendirme yapma konusunda yeterli görmemesi veya değerlendirme işinin sadece öğretmen tarafından yapılması gerektiğine inanması bu faktörler arasında sayılabilir. Bu durumda öz değerlendirmenin sürekli yapılması, öğrencilere öz değerlendirmeyi nasıl yapabileceklerinin net bir biçimde açıklanması ve öğrencilerin öz değerlendirme konusunda teşvik edilmesi söz konusu faktörlerin ortadan kaldırılmasında etkili olacaktır (Alicı, 2010; Tekindal, 2014).

Akran değerlendirmede ise öğrenciler ilgili kriterleri benzer statüdeki diğer öğrencilerin çalışmalarına uygulayarak başkalarının performans veya ürün kalitesini değerlendirirler. Böylece farklı çalışmaları inceleyip eleştirilerde bulunarak birbirleriyle ve birbirlerinden yeni bilgiler öğrenmiş olurlar. Akran değerlendirme, öğrencilere kendi çalışmalarının kalitesi hakkında akranlarından dönüt sağlamayı içerir. Akran dönütü, birlikte etkin olarak çalışmayı ve birlikte öğrenmeyi teşvik eder. Öğrenciler birbirlerinin güçlü ve zayıf yönlerini görerek öğrenme ihtiyaçları konusunda farkındalıklarını artırır. Hatta birbirlerini öğretmenlerinden daha iyi tanıyabilir ve daha ayrıntılı geri bildirim verebilirler. Bu açıdan akran değerlendirme ile kalabalık sınıflarda öğretmenin ulaşabileceğinden çok daha fazla sayıda öğrenciye geri bildirim verilmesi sağlanabilir. Böylece birbirlerinin yetenek ve becerilerini geliştirebilirler. Ancak öğrencilerin etkili bir akran değerlendirmesi yapma konusunda ustalaşmaları çok sayıda uygulama yapmalarını gerektirir. Ayrıca değerlendirme kriterleri de açık, uygun ve öğrencilerle tartışılarak belirlenmiş olmalıdır (Ballantyne ve diğerleri, 2002; Falchikov, 1986, 2001; Tekindal, 2014; Topping ve diğerleri, 2000).

Akran değerlendirme, hayatımızın ilk yıllarından itibaren başarı gelişimimizin bir parçası haline gelmektedir. Akranlarından aldıkları informal dönütler çocukların sosyal gelişimine önemli katkılarda bulunmaktadır. Akran dönütünün geliştirici gücü planlı değerlendirme etkinliklerine dahil edildiğinde öğrencilerin sosyal gelişimi önemli ölçüde hızlanabilir. Öğrenciler, iş birliği içinde takım çalışması yaparak ürünlerinin kalitesini artırma fırsatı bulurlar. Oluşturdukları ürünlerde fark edemedikleri hata ve eksiklikleri akranlarının bakış açısından görebilirler. Böylece ürünlerdeki aksaklıklar düzeltilerek çalışmalar daha üst seviyelere taşınabilir. Elbette böyle bir süreçte öğrencilerin ürünlerinin yanında; iletişim, iş birliği, tartışma gibi sosyal becerileri de gelişir. Öğrenciler birbirlerini yapıcı yönde eleştirmeyi ve eleştirileri hoşgörüle kabul etmeyi öğrenirler. Bu şekilde birlikte çalıştıklarında kendilerini topluluğun bir üyesi olarak görüp aidiyet duygusu geliştirebilirler. Akranlarından öğrendiklerini hem kendi kişisel gelişimlerinde hem de toplumun kişisel gelişiminde kullanabilen bireyler olarak yetişirler (Alicı, 2010; Tekindal, 2014).

Başlangıçta öz değerlendirme gibi akran değerlendirmenin de nesnel bir biçimde gerçekleştirilmesinde zorluklar yaşanabilir. Öğrencilerin kendilerine yakın buldukları ve sınıfça daha çok sevilen akranlarını değerlendirirken öznel davranma olasılıkları yüksektir. Ancak bu çalışmalar belirli aralıklarla rutin olarak gerçekleştirildiğinde, öğrencilerin daha başarılı değerlendirmeler yapmaya başladıkları görülecektir. Bu sürece yardımcı olmak için akran değerlendirmenin amacı, önemi ve uygulama adımları öğrencilere açıkça anlatılmalıdır. Değerlendirilen öğrencilerin başka özelliklerinin değerlendirilen özelliklerden ayırt edilmesinin gerekliliği vurgulanmalıdır. Öğrenciler yarıştırdıklarını değil birlikte çalıştıklarını hissetmeye başladıklarında değerlendirmelerin nesnellik düzeyi artacaktır. Ayrıca mümkün olan durumlarda değerlendirilen ürünün kime ait olduğunun bilinmemesi, bir çalışmayı birden fazla değerlendiricinin değerlendirmesi, değerlendiricilerin rastgele seçilmesi gibi uygulamalarda daha nesnel değerlendirmelere ulaşılmasına yardımcı olur (Bahar ve diğerleri, 2008; Alicı, 2010).

Öz değerlendirme ve akran değerlendirme birlikte kullanıldıklarında birbirlerine yardımcı olmakta ve birbirlerini geliştirmektedirler. Ancak bu gelişimin sağlanabilmesi için öğrencilerin değerlendirme becerilerini aktif ve doğru bir biçimde kullanabilmeleri gerekir. Bu ise öğrencilerin ölçme değerlendirme kültürü ile yetişmelerine fırsat tanınmasına bağlıdır. Yapılan araştırmalar, bu tür etkinliklerin ilköğretim birinci kademedden itibaren rutin bir biçimde gerçekleştirilmesinin, eleştirel düşünme becerilerine önemli katkıları olduğunu ortaya koymaktadır (Alicı, 2010). Ayrıca ilgili

alanyazın incelendiđinde bu tür deđerlendirmelerin sađlıklı bir şekilde yapılabilmesi için sürekli uygulanması gerektiđi görölmektedir. Bu nedenle öğrencilerin eğitimin ilk basamaklarından itibaren deđerlendirme etkinliklerine katılımları sađlanmalıdır. Öğrenci ve öğretmenin bakış açılarına birlikte yer veren deđerlendirme etkinlikleri daha detaylı verilere ulaşılmasını ve daha etkili çözüm yolları geliştirilmesini sađlayacaktır. Öğrencilerin bakış açısının öğretmenlerin bakış açısıyla karşılaştırılarak incelenmesi ise bu alandaki deđerlendirme etkinliklerinin geliştirilmesinde yol gösterici olacaktır. Bu durumda eğitimin ilk basamaklarında bulunan ilkokul öğrencilerinin, kendilerinin ve akranlarının çalışmalarını belirli kriterlere göre deđerlendirme konusunda alanın uzmanı olan öğretmenlerle farklılık gösterip göstermediklerini, gösteriyorlarsa bu farklılıđın düzeyini saptamak; deđerlendirme alanında bulunduđumuz noktayı belirleyebilmek açısından önem taşımaktadır. İlgili alanyazın incelendiđinde Türkiye’de öz ve akran deđerlendirme ile ilgili çok sayıda araştırma göröldüđü halde bu deđerlendirmelerin güvenilirliklerinin karşılaştırma yoluyla incelendiđi sınırlı sayıda çalışmaya rastlanmıştır. Bu tür deđerlendirmelerin önemli sınırlılıklarından birinin güvenilirlik olduđu göz önüne alındıđında bu konunun incelenmesinin öğrenci katımlı deđerlendirme çalışmalarında gelen seviyeyi ortaya koyabilmek açısından önemli olduđu düşünölmektedir. Bu nedenle; bu çalışmada ilkokul düzeyinde öğretmen ve öğrenci (öz ve akran) deđerlendirmelerinden elde edilen puanların güvenilirliklerinin belirlenmesi amaçlanmıştır. Bu amaç çerçevesinde ilkokul üçüncü sınıf düzeyinde Türkçe, Hayat Bilgisi ve Müzik derslerinde gerçekleştirilen örnek olaya dayalı performans görevlerinin öğretmen, öz ve akran deđerlendirmelerine ilişkin puanların güvenilirliklerinin nasıl deđiştirdiği incelenmiştir. Araştırmanın ilkokul öğrencilerinin öz deđerlendirme ve akran deđerlendirme becerilerinin incelenmesi yoluyla daha verimli deđerlendirme çalışmaları yapılmasına katkı sađlayacağı düşünölmektedir.

YÖNTEM

Bu çalışma, ilkokul üçüncü sınıf düzeyinde Türkçe, Hayat Bilgisi ve Müzik derslerinde uygulanan performans görevlerine ait yapılan öğretmen, öz ve akran deđerlendirmelerin güvenirliliđinin belirlenmesi amaçlandıđından betimsel çalışma özelliđini taşımaktadır.

Çalışma Grubu

Araştırmanın çalışma grubunu, Ankara ilinde yer alan bir ilkokulda öğrenim görmekte olan 14’ü erkek 16’sı kız olmak üzere toplam 30 üçüncü sınıf öğrencisi oluşturmaktadır. 30 öğrenci arasından tesadüfi olarak seçilen 5 öğrencinin akran deđerlendirme için puanlama yapmasına karar verilmiştir. Sonuç olarak; geriye kalan 25 öğrenci çalışmanın ölçme objesi olarak çalışmada yer almıştır.

Veri Toplama Araçları

Veri toplama aracı olarak Millî Eğitim Bakanlığı tarafından hazırlanan ve öğretmen kılavuz kitaplarında yer alan deđerlendirme, öz deđerlendirme ve akran deđerlendirme ölçekleri, araştırmanın amacına uygun şekilde konu alanı uzman görüşleri (5 Sınıf öğretmeni, 2 Türkçe öğretmeni ve 1 Müzik öğretmeni) dođrultusunda hazırlanan performans görevlerine uygun olarak sadeleştirilerek kullanılmıştır.

Türkçe dersinde yazma becerilerinin ölçülmesi amacıyla, Millî Eğitim Bakanlığı tarafından 2013-2014 öğretim yılından itibaren kullanılması uygun görölen öğretmen kılavuz kitabında yer alan Yazma Becerileri Deđerlendirme Ölçeđi kullanılmıştır (Millî Eğitim Bakanlığı-MEB, 2013). Uygulamanın süresi ve performans görevi göz önüne alınarak ölçekte yer alan 10 ölçüt arasından, uzman görüşleri dođrultusunda kritik ölçütlerin seçilip düzenlenmesiyle dört ölçütün yer aldığı puanlama anahtarı oluşturulmuştur. Oluşturulan puanlama anahtarı Tablo 1’de verilmiştir.

Tablo 1. Türkçe Performans Görevinde Kullanılan Puanlama Anahtarı

Ölçüt	Derecelendirme		
	1	2	3
Yazım kurallarına uyma			
Anlamli ve kurallı cümleler yazma			
Olayları oluş sırasına göre yazma			
Yazılarında ana düşünceye yer verme			

Hayat Bilgisi dersinde tartışma becerilerinin ölçülmesi amacıyla, Millî Eğitim Bakanlığı tarafından 2015-2016 öğretim yılından itibaren kullanılması uygun görülen öğretmen kılavuz kitabında yer alan Tartışma Ölçeği kullanılmıştır (MEB, 2017a). Uygulamanın süresi ve hazırlanan performans görevi göz önüne alınarak ölçekte yer alan 10 ölçüt arasından, uzman görüşleri doğrultusunda kritik ölçütlerin seçilip düzenlenmesiyle dört ölçütten oluşan puanlama anahtarı hazırlanmıştır. Oluşturulan puanlama anahtarı Tablo 2’de verilmiştir.

Tablo 2. Hayat Bilgisi Performans Görevinde Kullanılan Puanlama Anahtarı

Ölçüt	Derecelendirme		
	1	2	3
Fikrini açıkça ortaya koyabilme, söyleyebilme			
Soruları doğru yorumlama ve sorulara uygun cevaplar verme			
Tartışma kurallarına uyma			
Ses tonu ve vücut hareketlerini kontrol etme			

Müzik dersinde şarkı söyleme performansının ölçülmesi amacıyla, Millî Eğitim Bakanlığı tarafından 2017-2018 öğretim yılından itibaren kullanılması uygun görülen öğretmen kılavuz kitabında yer alan Şarkı/Türkü/Marş Performansına Yönelik Analitik Dereceli Puanlama Anahtarı kullanılmıştır (MEB, 2017b). Uygulamanın süresi ve performans görevi göz önüne alınarak puanlama anahtarında yer alan altı ölçüt arasından, uzman görüşleri doğrultusunda kritik ölçütlerin seçilip düzenlenmesiyle dört ölçütten oluşan puanlama anahtarı oluşturulmuştur. Oluşturulan puanlama anahtarı Tablo 3’te verilmiştir.

Tablo 3. Müzik Performans Görevinde Kullanılan Puanlama Anahtarı

Ölçüt	Derecelendirme		
	1	2	3
Şarkının sözlerini doğru söyleme			
Şarkının ezgisini doğru söyleme			
Şarkının ritmine dikkat etme			
Sesini doğru ve etkili kullanma			

Sonuç olarak; her bir derste dört ölçütten oluşan ve her birinin ölçütü derse özgü olan puanlama anahtarları kullanılmıştır. Puanlama anahtarları yaş grubuna uygun olarak üç farklı gülen surat simgesi ile derecelendirilmektedir. Gülen surat simgelerinin puan olarak karşılığı 1-2-3 olarak araştırmacılar tarafından veri seti hazırlanmıştır.

İşlem

Araştırmada, disiplinler arası yaklaşım kullanılmıştır. Değerler eğitimi kapsamındaki yardımseverlik konusu, üç derse yönelik uygulamalarla ilişkilendirilerek ele alınmıştır. Bu yaklaşımla aynı konunun farklı disiplinlerin yöntemleri açısından incelenmesinin mümkün olacağı ve bu şekilde daha detaylı veriler elde edileceği düşünülmüştür. Ayrıca konuyla ilgili farklı bilgi ve becerilerin organize edilerek anlamlı bir bütün olarak kavratılmasıyla öğrencilerin daha sağlıklı bir değerlendirme yapabilmesi

amaçlanmıştır. Aynı zamanda öđretmen ve öđrencilerin farklı disiplinlerin deđerlendirilmesi konusundaki bakış açısı farklılıklarının da incelenmesinin mümkün olacağı düşünölmüştür.

Araştırmada aynı öđrenci grubunun üç farklı ders için hem akran hem de öz deđerlendirme yapmaları istenilmiştir. Öđrenci grubu aynı kalırken, farklı derslerde farklı öđretmenlerin deđerlendirme yapmaları sağlanmıştır. Bu durumda araştırma başlamadan önce ilk olarak öđretmen ve öđrenciler deđerlendirme türleri hakkında bilgilendirilmiştir. Öđretmen deđerlendirmeleri, her biri alanında uzman ve deneyime sahip devlet okulunda çalışan iki sınıf öđretmeni ve bir müzik öđretmeni tarafından yapılmıştır. Hayat Bilgisi dersinde araştırma yapılan sınıfın öđretmeni deđerlendirme yaparken, Türkçe dersinde farklı bir sınıf öđretmeni ve Müzik dersinde de araştırmacılar arasında yer alan Müzik öđretmeni tarafından deđerlendirme yapılmıştır. Müzik dersinin ayrı becerileri gerektirmesinden ve puanlamanın olabildiğince yansız yapılmasından dolayı Müzik dersinin Müzik branşı eğitimi almış bir öđretmen tarafından yapılmasına karar verilmiştir. Müzik öđretmeni araştırmanın içinde yer aldığı için puanlama anahtarı ve puanlama süreci hakkında detaylı bilgiye sahiptir. Türkçe ve Hayat Bilgisi derslerinin puanlamasını yapan sınıf öđretmenlerine hem deđerlendirme türleri hem de puanlama anahtarları hakkında çalışma öncesinde bilgilendirme yapılmıştır. Bunun için sınıf öđretmenlerine, puanlama anahtarı kullanarak deđerlendirmenin nasıl yapılacağı hakkında eğitim verilmiş ve öđretmenlerle birlikte uygulama örnekleri incelenmiştir. Öđrencilerle yapılan bilgilendirme sürecinde de öđretmen deđerlendirmesi, öz deđerlendirme ve akran deđerlendirmesi ve puanlama anahtarları hakkında kısa bir eğitim verilmiştir.

Türkçe dersinde öđrencilere, yardımseverliđin önemini konu alan bir masalın anlatıldığı çizgi film izletilmiştir. Masal yarısında durdurularak öđrencilerden masalın sonunu tamamlamaları istenmiştir. Tüm öđrenciler çalışmalarını tamamladıktan sonra, sırayla tahtaya kalkarak masalın tamamladıkları kısmını okumuşlardır. Yazım kuralları da deđerlendirme kapsamına alındığından okumayı bitiren öđrencinin kâğıdı, akran deđerlendirmesi yapan öđrenciler ve öđretmen tarafından incelenmiştir. Bu şekilde öđrencilerin yazma becerileri, öđretmen ve öđrenciler tarafından deđerlendirilmiştir.

Hayat Bilgisi dersinde öđrencilere, yapılan bir yardımın insanlardaki yardımseverlik duygusunu uyandırarak nasıl bir döngü oluşturabileceđini anlatan kısa bir film izletilmiştir. Ardından öđrencilerden, izledikleri olaylara dayalı olarak yardımseverliđin ortaya çıkarabileceđi olumlu ve olumsuz sonuçları gruplar halinde tartışmaları istenmiştir. Öđrencilerin isteklerine göre dörder kişilik gruplar oluşturulmuş ve karşılıklı iki grup halinde konuyu tartışmaları sağlanmıştır. Her grup tartışmasını bitirdikten sonra öđrencilerin grup içerisindeki tartışma becerileri; öđretmen ve öđrenciler tarafından hazırlanan puanlama anahtarları kullanılarak deđerlendirilmiştir.

Müzik dersinde öđrencilere, yardımseverliđin öneminin anlatıldığı bir şarkı, kulaktan öđretim yöntemi kullanılarak öđretilmiştir. Daha sonra öđrencilerden şarkıyı bireysel olarak seslendirmeleri istenmiştir. Öđrencilerin gösterdiği şarkı performansı öđretmen ve öđrenciler tarafından deđerlendirilmiştir.

Verilerin Analizi

Bu çalışmada; öđretmen, öz ve akran deđerlendirme sonucu verilen puanların güvenirliliđinin belirlenmesi amaçlanmaktadır. Alanyazın incelendiğinde; ölçme sonuçlarına ait güvenirliliđin belirlenmesinde Klasik Test Kuramı (KTK), Genelenebilirlik (G) Kuramı ve Madde Tepki Kuramının (MTK) işe koşulduđu görölmektedir (Güler, 2011). Özellikle farklı puanlayıcılar arasındaki güvenirliliđin belirlenmeye çalışıldığı çalışmalara odaklanıldığında, son zamanlarda G Kuramı ya da MTK'ye dayalı yöntemlerin KTK'ye nazaran daha sık tercih edildiđi görölmektedir (Atılğan, 2005; Börkan, 2017; Büyükkıdık & Anıl, 2015; Farrokhi, Esfandiari & Dalili, 2011; Farrokhi, Esfandiari & Schaefer, 2012; Karakaya, 2015; Matsuno, 2009; Nalbantođlu-Yılmaz, 2017; Taşdelen-Teker, Şahin & Baytemir, 2016; Yıldıztekin, 2014). KTK ve G Kuramı temelinde bir karşılaştırma yapılırsa; KTK temeline dayalı yapılan güvenirlilik belirleme çalışmalarında yalnızca bir hata kaynađına bađlı kestirim yapılmasına izin verilirken, G kuramına dayalı yapılan güvenirlilik belirleme çalışmalarında tüm hata kaynaklarının analize dahil edilebildiđi görölmektedir. Buna ek olarak; G Kuramında hata kaynakları ayrı ayrı ele alınabildiđi gibi, hata kaynaklarının etkileşimleri de analiz sonucunda

belirlenebilmektedir (Brennan, 2001; Güler, 2009; Güler, Kaya-Uyanık & Taşdelen-Teker, 2012; Shavelson & Webb, 1991). Taşdelen-Teker ve Güler'in (2019) yaptığı çalışmaya bakıldığında da G Kuramının özellikle puanlayıcılar arası güvenilirlik ve standart belirleme çalışmalarında sıklıkla kullanıldığı görülmektedir. G Kuramının bu üstünlüklerinden ve kullanım alanından dolayı; farklı puanlayıcı türleri (öğretmen-öz-akran) arasındaki güvenilirliğin belirlenmesinin amaçlandığı bu çalışmada G Kuramı tercih edilmiştir.

Çalışmanın ölçme objesi olan öğrenciler (s) ile puanlayıcı türü (r) ve ölçüt (c) değişkenlik kaynakları kullanılarak çalışmacının amacı doğrultusunda tümüyle çaprazlanmış iki yüzeyli tesadüfi desen (srxrc) üzerinden analizler yürütülmüştür. Puanlayıcı türlerinden biri olan akran değerlendirme analize dahil edilirken beş öğrencinin 25 öğrenciye vermiş oldukları puanların ortalaması alınmıştır. Çalışma kapsamında ele alınan üç ders için de değişkenlik kaynaklarını oluşturan değişkenlerin ana ve ortak etkilerinin belirlenmesi için kestirilen varyans bileşenleri, G ve Phi katsayıları hesaplanmıştır. Ayrıca G-Yüzeyleri analizi kullanılarak puanlayıcı türlerinin sırasıyla analizden çıkarılması durumunda elde edilen G ve Phi katsayıları da hesaplanmıştır. Analizler EduG 6.1 paket programı kullanılarak gerçekleştirilmiştir.

BULGULAR

Bu bölümde öğretmen, öz ve akran değerlendirmede yapılan puanlamaların kestirilen varyans bileşenleri, güvenilirlik değerleri ve puanlayıcı türüne göre G-Yüzey bileşenleri sırasıyla Türkçe, Hayat Bilgisi ve Müzik dersleri için ayrı başlıklarda verilmiştir.

1. Türkçe Dersi

Türkçe dersinde tümüyle çaprazlanmış olan srxrc desenine ait G çalışması için; kestirilen varyans bileşenleri ve toplam varyans açıklama yüzdeleri s, r ve c ana etkileri ile sr, sc, rc ve src ortak etkileri Tablo 4'te verilmiştir.

Tablo 4. Türkçe Dersine İlişkin Kestirilen Varyans Bileşenleri

Varyans Kaynağı	Toplam Kareler	Sd	Kareler Ortalaması	Varyans (σ^2)	%
Öğrenci (s)	77.16853	24	3.21536	0.25745	64.0
Puanlayıcı Türü (r)	4.10027	2	2.05013	0.01712	4.3
Ölçüt (c)	3.32627	3	1.10876	0.01068	2.7
sr	6.17307	48	0.12861	0.00674	1.7
sc	7.12373	72	0.09894	-0.00090	0.0
rc	1.86453	6	0.31076	0.00836	2.1
src,e	14.63547	144	0.10164	0.10164	25.3
Toplam	114.39187	299			100%

Tablo 4'e bakıldığında, Türkçe dersinde öğrenci (s) ana etkisi için kestirilen varyans bileşeninin (0.258) toplam varyansın %64'ünü açıkladığı görülmektedir. Öğrenci ana etkisi toplam varyans içerisinde en büyük paya sahip olmaktadır. Böylece yapılan ölçme işlemi ile öğrenciler arası farklılıkların belirlenebildiği sonucu elde edilmektedir.

Puanlayıcı türü (r) ana etkisi için kestirilen varyans bileşeninin (0.017) toplam varyansın %4.3'ünü açıkladığı görülmektedir. Puanlayıcı türü ana etkisi ise toplam varyans içerisinde en büyük üçüncü paya sahip olan varyans bileşeni olmuştur. Buna göre; öğretmen, öz ve akran değerlendirmeleriyle verilen puanların bir miktar farklılık gösterdiği söylenebilir.

Ölçüt (c) ana etkisi için kestirilen varyans bileşeninin (0.011) toplam varyansın %2.7'sini açıkladığı görülmektedir. Bu durumda, verilen puanların bir ölçütten diğer ölçüte az da olsa farklılık gösterdiği söylenebilir.

Ortak etki deđerleri incelendiđinde; öğrenci-puanlayıcı türü (sr) ortak etkisi için kestirilen varyans bileşeninin (0.007) toplam varyansın %1.7'sinin açıkladıđı görölmektedir. Öğrenci-puanlayıcı türü (sr) ortak etkisinin toplam varyans içerisinde en düşük ikinci varyansa sahip olmuştur. Bu durumda, öğrencilere farklı puanlayıcı türlerinin verdiđi puanların çok deđişmediđi söylenebilir.

Öğrenci-ölçüt (sc) ortak etkisi için kestirilen varyans bileşeninin (-0.001) toplam varyansın %0.0'ını açıkladıđı görölmektedir. Öğrenci-ölçüt (sc) ortak etkisinin ise toplam varyans içerisinde en düşük varyansa sahip olurken negatif deđer almıştır. Varyansın negatif deđer aldıđı durumlarda, Cronbach ve diđerleri (1972) varyans deđerinin sıfır alınmasını önermişlerdir (akt., Dođan & Anadolu, 2017). Varyansın negatif çıkma sebebi; çalışma grubunun küçük olması ya da ölçme deseninin uygun olmaması olabilir (Taşdelen-Teker ve diđerleri, 2016). Bu çalışmada; desenle ilgili bir sorun görülmediđi için bulgunun çalışma grubunun büyüklüğüyle ilgili olduđu düşünölmektedir. Buna dayanarak öğrenci-ölçüt (sc) ortak etkisinin toplam varyansı sıfır olarak ele alındıđında, bu etkinin toplam varyansa herhangi bir katkı yapmadıđı söylenebilir. Kısacası, öğrencilerin performansları ölçütlere göre farklılık göstermemiştir.

Puanlayıcı türü-ölçüt (rc) ortak etkisi için kestirilen varyans bileşeninin (0.008) toplam varyansın %2.1'inin açıkladıđı görölmektedir. Bu bulgu, puanlayıcı türüne göre bir ölçütten diđer ölçüte yapılan puanlamalarda az da olsa bir farklılıđın olduđunu göstermektedir.

Öğrenci-puanlayıcı türü-ölçüt (artık) ortak etkisi varyans bileşeni (0.102) toplam varyansın %25.3'ünü açıkladıđı görölmektedir. Bu oran toplam varyans içindeki en büyük ikinci deđer olma özelliđine sahiptir. Oysa ki, öğrenci-puanlayıcı türü-ölçüt (artık) ortak etkisi varyans bileşeninin toplam varyans içindeki payının küçük olması beklenmektedir (Shavelson & Webb, 1991). Sonuç olarak bu durum; öğrenci-puanlayıcı türü-ölçüt ortak etkisinin ve/veya ölçmedeki tesadüfi hataların büyük olabileceđinin göstergesi olabilir.

G ve Phi katsayılarına bakıldıđında; göreli hata varyansına dayalı olarak hesaplanan G katsayısı .96, mutlak hata varyansına dotalı olarak hesaplanan Phi katsayısı ise .93 olarak elde edilmiştir. Bu deđerlerin güvenilirlik katsayısının kabul edilebilir sınırları içerisinde oldukça yüksek deđerler olduđu söylenebilir (Brennan, 2001).

G-yüzeyleri analizi sonucunda puanlayıcı türlerinden her birinin sırasıyla analize dahil edilmediđinde elde edilen güvenilirlik katsayıları Tablo 5'te verilmiştir.

Tablo 5. Puanlayıcı Türlerinin G-Yüzey Analizi

Yüzey	Düzey	G	Φ
Puanlayıcı Türü (n _r = 3)	Öğretmen Deđerlendirmesi	.92	.88
	Öz Deđerlendirme	.97	.94
	Akran Deđerlendirmesi	.92	.86

Tablo 5'e bakıldıđında; analizden öğretmen ya da akran deđerlendirmelerinin çıkarılması sonucunda hesaplanan G ve Φ katsayılarında az da olsa bir düşme olmaktadır. Ancak yine de elde edilen güvenilirlik katsayıları oldukça yüksektir. Analizden öz deđerlendirmenin çıkarılması sonucunda ise hem G hem de Φ katsayısının yükseldiđi görölmektedir.

2. Hayat Bilgisi Dersi

Hayat Bilgisi dersinde tümüyle çaprazlanmış olan srxrc desenine ait G çalışması için; kestirilen varyans bileşenleri ve toplam varyans açıklama yüzdeleri s, r ve ana etkileri ile sr, sc, rc ve src ortak etkileri Tablo 6'da verilmiştir.

Tablo 6. Hayat Bilgisi Dersine İlişkin Kestirilen Varyans Bileşenleri

Varyans Kaynağı	Toplam Kareler	Sd	Kareler Ortalaması	Varyans (σ^2)	%
Öğrenci (s)	109.41813	24	4.55909	0.35791	74.6
Puanlayıcı Türü (r)	1.01840	2	0.50920	0.00034	0.1
Ölçüt (c)	0.59987	3	0.19996	-0.00254	0.0
sr	9.62827	48	0.20059	0.03713	7.7
sc	8.33013	72	0.11570	0.02120	4.4
rc	1.95973	6	0.32662	0.01098	2.3
src,e	7.50027	144	0.05209	0.05209	10.9
Toplam	138.45480	299			100%

Tablo 6'ya bakıldığında; Hayat Bilgisi dersinde öğrenci (s) ana etkisi için kestirilen varyans bileşeninin (0.358) toplam varyansın %74.6'sını açıkladığı görülmektedir. En yüksek varyans oranının öğrenci değişkeninden elde edilmesi sonucunda, gerçekleştirilen ölçme işlemi ile öğrenciler arası farklılıkların belirlendiği elde edilmiştir.

Puanlayıcı türü (r) ana etkisi için kestirilen varyans bileşeninin (0.000) toplam varyansın %0.1'ini açıkladığı görülmektedir. Puanlayıcı türü ana etkisi ise toplam varyans içerisinde en küçük ikinci paya sahip olan varyans bileşeni olmuştur. Buna göre; öğretmen, öz ve akran değerlendirmelerinde verilen puanların neredeyse hiç anlamlı bir farklılık göstermediği söylenebilir.

Ölçüt (c) ana etkisi için kestirilen varyans bileşeninin (-0.003) toplam varyansın %0.0'ını açıkladığı görülmektedir. Ölçüt ana etkisinin ise toplam varyans içerisinde en düşük varyansa sahip olurken negatif değer almıştır. Bu değişkenin toplam varyansı sıfır olarak ele alınırsa, bu etkinin toplam varyansa herhangi bir katkı yapmadığı söylenebilir. Kısacası, ölçütlere göre puanlamalar farklılık göstermemiştir.

Ortak etki değerleri incelendiğinde; öğrenci-puanlayıcı türü (sr) ortak etkisi için kestirilen varyans bileşeninin (0.037) toplam varyansın %7.7'sinin açıkladığı görülmektedir. Öğrenci-puanlayıcı türü (sr) ortak etkisinin toplam varyans içerisinde en yüksek üçüncü varyansa sahip olmuştur. Bu durumda, öğrencilere farklı puanlayıcı türlerinin verdiği puanların farklılık gösterdiği söylenebilir.

Öğrenci-ölçüt (sc) ortak etkisi için kestirilen varyans bileşeninin (0.021) toplam varyansın %4.4'ünün açıkladığı görülmektedir. Öğrenci-ölçüt (sc) ortak etkisinin ise toplam varyans içerisinde en düşük dördüncü varyansa sahip olurken olmuştur. Sonuç olarak, öğrencilerin performansları ölçütlere göre az da olsa farklılık göstermektedir.

Puanlayıcı türü-ölçüt (rc) ortak etkisi için kestirilen varyans bileşeninin (0.011) toplam varyansın %2.3'ünün açıkladığı görülmektedir. Bu durum, puanlayıcı-ölçüt (rc) ortak etkisinin en düşük üçüncü varyans değerine sahip olduğunu gösterirken, puanlayıcı türünün ölçütten ölçüte çok az farklılaştığı söylenebilir.

Öğrenci-puanlayıcı türü-ölçüt (artık) ortak etkisi varyans bileşeni (0.053) toplam varyansın %10.9'unu açıkladığı görülmektedir. Bu oranın toplam varyans içindeki en büyük ikinci değere sahip olduğu görülürken, öğrenci-puanlayıcı türü-ölçüt ortak etkisinin ve/veya ölçmedeki tesadüf hataların büyük olabileceğinin bir göstergesi olabilir.

G ve Phi katsayıları incelendiğinde hem G katsayısının hem de Phi katsayısının ise .94 olarak elde edilmiştir. Elde edilen göreceli ve mutlak güvenilirlik katsayılarının kabul edilebilir sınırları içerisinde oldukça yüksek değerler olduğu söylenebilir.

G-yüzeyleri analizi sonucunda puanlayıcı türlerinden her birinin sırasıyla analize dahil edilmediğinde elde edilen güvenilirlik katsayıları Tablo 7'de verilmiştir.

Table 7. Puanlayıcı Türlerinin G-Yüzey Analizi

Yüzey	Düzye	G	Φ
Puanlayıcı Türü ($n_r = 3$)	Öğretmen Değerlendirmesi	.91	.90
	Öz Değerlendirme	.97	.96
	Akran Değerlendirmesi	.89	.88

Tablo 7'ye bakıldığında; analizden öğretmen ya da akran deđerlendirmelerinin çıkarılması sonucunda hesaplanan G ve Φ katsayılarında az da olsa bir düşüş olduđu görölmektedir. Bu düşüşün akran deđerlendirmede az da olsa daha fazla olduđu, ancak yine de elde edilen güvenirlilik katsayılarının oldukça yüksek olduđu görölmektedir. Öz deđerlendirmenin analizden çıkarıldığı G ve Φ katsayılarına bakıldığında ise her iki güvenirlilik katsayısının da bir miktar yükseldiđi görölmektedir.

3. Müzik Dersi

Müzik dersinde tümüyle çaprazlanmış olan srxrc desenine ait G çalışması için; kestirilen varyans bileşenleri ve toplam varyans açıklama yüzdeleri s, r ve ana etkileri ile sr, sc, rc ve src ortak etkileri Tablo 8'de verilmiştir.

Tablo 8. Müzik Dersine İlişkin Kestirilen Varyans Bileşenleri

Varyans Kaynađı	Toplam Kareler	Sd	Kareler Ortalaması	Varyans (σ^2)	%
Öđrenci (s)	71.25013	24	2.96876	0.21139	47.3
Puanlayıcı Türü (r)	5.35707	2	2.67853	0.02104	4.7
Ölçüt (c)	0.18600	3	0.06200	-0.00453	0.0
sr	17.49627	48	0.36451	0.06021	13.5
sc	13.77067	72	0.19126	0.02253	5.0
rc	2.00400	6	0.33400	0.00841	1.9
src,e	17.80933	144	0.12368	0.12368	27.7
Toplam	127.87347	299			100%

Tablo 8'e bakıldığında; Müzik dersinde öğrenci (s) ana etkisi için kestirilen varyans bileşeninin (0.211) toplam varyansın %47.3'ünü açıkladığı görölmektedir. Müzik dersi kapsamında gerçekleştirilen puanlama sonucunda öğrenciler arası farklılıkların belirlenebildiđi sonucu elde edilmektedir.

Puanlayıcı türü (r) ana etkisi için kestirilen varyans bileşeninin (0.021) toplam varyansın %4.7'sini açıkladığı görölmektedir. Puanlayıcı türü ana etkisine bakıldığında; öğretmen, öz ve akran deđerlendirmelerinde verilen puanların farklılık gösterdiđi söylenebilir.

Ölçüt (c) ana etkisi için kestirilen varyans bileşeninin (-0.05) toplam varyansın %0.0'ını açıkladığı görölmektedir. Ölçüt ana etkisinin ise toplam varyans içerisinde en düşük varyansa sahip olurken negatif deđer almıştır. Bu deđerşkenin toplam varyansı sıfır olarak ele alınırsa, bu etkinin toplam varyansa herhangi bir katkı yapmadığı söylenebilir. Kısacası, ölçütlere göre puanlamalar farklılık göstermemiştir.

Ortak etki deđerleri incelendiğinde; öğrenci-puanlayıcı türü (sr) ortak etkisi için kestirilen varyans bileşeninin (0.060) toplam varyansın %13.5'inin açıklandığı görölmektedir. Bu durumda; öğrenci-puanlayıcı türü (sr) ortak etkisinin toplam varyans içerisinde en yüksek üçüncü varyansa sahip olurken, elde edilen bu bulgu ile öğrencilere farklı puanlayıcı türlerinin verdiđi puanların farklılaştığı söylenebilir.

Öđrenci-ölçüt (sc) ortak etkisi için kestirilen varyans bileşeninin (0.023) toplam varyansın %5.0'ini açıklarken, buna göre; öğrencilere ölçütlere göre verilen puanların farklılaştığı söylenebilir. Puanlayıcı türü-ölçüt (rc) ortak etkisi için kestirilen varyans bileşenine bakıldığında ise; toplam varyansın %1.9'unu açıkladığı görölmektedir. Bu sonuca göre; çok az da olsa puanlayıcı türünün ölçütlere göre verdiđi puanlar farklılaşmaktadır.

Öđrenci-puanlayıcı türü-ölçüt (artık) ortak etkisi varyans bileşeni (0.124) toplam varyansın %27.7'sini açıklarken, elde edilen bu deđer toplam varyans içindeki en büyük ikinci deđer olma özelliđine sahip olmuştur. Bu sebeple, öğrenci-puanlayıcı türü-ölçüt ortak etkisinin ve/veya ölçmedeki tesadüfi hataların büyük olabileceđi söylenebilir.

Analiz sonucunda elde edilen G ve Phi katsayılarına bakıldığında; G katsayısının .85, Phi katsayısının da .83 olduğu görülmüştür. Elde edilen güvenilirlik katsayılarının literatüre göre kabul edilen sınırlar içerisinde yer aldığı görülmektedir (Brennan, 2001).

G-yüzeyleri analizi sonucunda puanlayıcı türlerinden her birinin sırasıyla analize dahil edilmediğinde elde edilen güvenilirlik katsayıları Tablo 9’da verilmiştir.

Tablo 9. Puanlayıcı Türlerinin G-Yüzey Analizi

Yüzey	Düzye	G	Φ
Puanlayıcı Türü (n _r = 3)	Öğretmen Değerlendirmesi	.63	.60
	Öz Değerlendirme	.97	.96
	Akran Değerlendirmesi	.68	.62

Tablo 9’a bakıldığında; öğretmen ya da akran değerlendirmelerinin analizden çıkarılmasıyla elde edilen G ve Φ katsayılarında ciddi bir düşme olduğu görülmektedir. Elde edilen güvenilirlik katsayılarının alanyazında belirtilen kabul edilebilir güvenilirlik katsayısı sınırından da düşük olduğu görülmektedir. Öz değerlendirme türünün analizden çıkarılması sonucunda ise hem G hem de Φ katsayısında bir yükselme olmuştur.

SONUÇLAR ve TARTIŞMA

Bu çalışmada ilkökul üçüncü sınıf öğrencilerinin, öz ve akran değerlendirme yöntemleriyle verdikleri puanlar ile öğretmen değerlendirmesi sonucunda verilen puanların güvenilirliğinin belirlenmesi amaçlanmıştır. Bu amaç doğrultusunda, disiplinlerarası yaklaşım ele alınarak değerler eğitimi kapsamındaki yardımseverlik konusu; Türkçe, Hayat Bilgisi ve Müzik derslerine yönelik uygulamalarla ilişkilendirilmiştir.

Birden fazla hata kaynağının analize dahil edilmesi amaçlandığı ve varyans kaynaklarının detaylı olarak incelenmesi istendiği için çalışmada G Kuramı kullanılmıştır. İlgili kuramın avantajları sayesinde; varyans kaynaklarının hem ana hem de etkileşimli etkisi incelenmiş olup görece ve mutlak güvenilirlik katsayıları kestirilmiştir.

Türkçe dersine bakıldığında; toplam varyansı en çok açıklayan bileşenin öğrenci (s) ana etkisi olduğu görülmektedir. En büyük açıklama yüzdesine öğrenci (s) ana etkisinin sahip olması ölçme işlemi açısından istenilen bir durum olma özelliği taşımaktadır. Çünkü yapılan ölçme işlemiyle öğrenciler arasındaki farklılıkların ortaya çıkarılabildiği sonucu elde edilmektedir (Atılğan, 2005; Doğan & Anadol, 2017; Taşdelen-Teker ve diğerleri, 2016). Toplam varyansı, öğrenci ana etkisinden sonra, en çok açıklayan bileşenin artıklar bileşeni olduğu görülmektedir. Elde edilen bu sonuç öğrenci-puanlayıcı türü-ölçüt (src,e) ortak etkisinin ve/veya tesadüfi hataların büyük olabileceğinin bir göstergesi olabilir. Bu ders kapsamında tesadüfi hataların sebebinin; bu tür uygulamalarla sık karşılaşmayan öğrencilerin yaşadıkları heyecan ve motivasyon eksikliği olduğu söylenebilir. Ölçüt (c) değişkeninin ana etkisine bakıldığında; toplam varyansın açıklanmasına %2.7 oranında katkı sağladığı görülmektedir. Ölçütler açısından değerlendirildiğinde; yazma becerileri kapsamındaki ölçütlerden bazılarında öğrenci ve öğretmen bakış açılarının farklılık gösterdiği söylenebilir. Türkçe dersi kapsamında elde edilen dikkat çekici bir diğer bulgu ise öğrenci-ölçüt (sc) ortak etkisinin toplam varyansa herhangi bir katkı yapmamış olmasıdır. Kısacası, öğrencilerin performansları puanlama anahtarında yer alan ölçütlere göre bir farklılık göstermemektedir. Bu durumda; söz konusu ölçütlerin aynı becerileri ölçtüğü söylenebilir.

Hayat Bilgisi dersi kapsamında sonuçlar ele alındığında; toplam varyansı en çok açıklayan bileşen öğrenci (s) ana etkisi olmuş, öğrenci ana etkisinden sonra toplam varyansın açıklanmasında en büyük paya ise artıklar (src,e) sahip olmuştur. Öğrenci ana etkisinin en büyük paya sahip olmasıyla Hayat Bilgisi dersi kapsamında yapılan değerlendirmelerde de öğrenciler arasındaki farklılıkların ortaya çıkarılabildiği söylenebilir. Bu ders kapsamında meydana gelebilecek tesadüfi hata kaynaklarının ise;

etkinlik sırasında etkinliğe katılmayan öğrenciler tarafından tartışmaya müdahale edilmesiyle oluşan dikkat dağınıklığı ve sınıfta oluşan gürültü olabileceği düşünülmektedir. Hayat Bilgisi dersine ilişkin kestirilen varyans değerlerinde puanlayıcı türü (r) ana etkisi toplam varyans içerisinde oldukça küçük bir paya sahip olmuştur. Bir başka ifadeyle öğretmen, öz ve akran değerlendirmelerinde verilen puanların neredeyse hiç anlamlı bir farklılık göstermediği söylenebilir. Araştırmada Hayat Bilgisi dersine ilişkin öğretmen değerlendirmesinin, araştırmanın yapıldığı sınıfın öğretmeni tarafından yapıldığı düşünüldüğünde elde edilen sonucun öğretmenin sınıftaki öğrencileri daha iyi tanımasından dolayı ve öğrencilerin sınıf öğretmenlerinin yaptığı bir değerlendirme ortamında daha rahat şekilde puanlama yapabileceklerinden kaynaklandığı düşünülebilir.

Müzik dersi kapsamında kestirilen varyans bileşenlerine ilişkin sonuçlar ele alındığında; öğrenci (s)ana etkisi, diğer derslerde olduğu gibi Müzik dersinde de toplam varyansı en çok açıklayan bileşen olmuştur. Bu açıdan Müzik dersi kapsamında yapılan değerlendirmelerde de öğrenciler arası farklılıklar ortaya çıkarılmıştır. Ayrıca artıkların (src,e) toplam varyansın açıklanmasındaki etki öğrenci (s) ana etkisinden sonraki en büyük paya sahip olmuştur. Müzik dersindeki artıkların yüksek bir paya sahip olmasının sebepleri arasında; bazı öğrencilerin bireysel performansı sırasında sınıftan gelen tepkilerin gürültüye yol açması ve bireysel performans göstermeye alışık olmayan öğrencilerin heyecanlanması gibi tesadüfi hata kaynaklarının karışmış olması sayılabilir. Müzik dersi kapsamında elde edilen bir diğer dikkat çekici bulgu; ölçüt (c) ana etkisinin toplam varyansa herhangi bir katkı yapmaması, başka bir deyişle ölçütlere göre puanlamaların farklılık göstermemesi olmuştur. Bu durum, ölçütlerin tümünün şarkı söyleme becerilerine yönelik olması ve öğrencilerin müzik yeteneği seviyesinin ölçütlerin ilgili olduğu beceriler üzerinde aynı seviyede etkili olması ile açıklanabilir.

G ve Φ katsayılarına ilişkin bulgulara bakıldığında her üç ders için de elde edilen G ve Φ katsayılarının alanyazında kabul gören .80 ölçütünden (Brennan, 2001) oldukça yüksek olduğu görülmektedir. G ve Φ katsayıları, dersler bazında ele alındığında ise; müzik dersinde elde edilen katsayıların, Türkçe ve Hayat Bilgisi derslerinde edilen katsayılardan düşük olduğu görülmektedir. Türkçe ve Hayat Bilgisi derslerinde elde edilen katsayılar ise .90'ın üzerinde olup birbirlerine oldukça yakın değerler almışlardır. Çalışmada Türkçe ve Hayat Bilgisi derslerindeki öğretmen değerlendirmelerinin sınıf öğretmenleri tarafından, Müzik dersindeki değerlendirmenin ise branş öğretmeni olarak Müzik öğretmeni tarafından yapılmış olması güvenirliliği etkileyen bir sebep olarak düşünülebilir.

Puanlayıcı türlerine ilişkin G-Yüzeyleri analizi sonucunda elde edilen değerlere bakıldığında; her üç ders kapsamında yapılan değerlendirmeler için öğretmen ve akran değerlendirmelerinin analize eklenmemesi durumunda G ve Φ katsayılarında düşüşler meydana gelmektedir. Bu düşüşler sonucunda elde edilen yeni G ve Φ katsayıları; Türkçe ve Hayat Bilgisi dersleri için kabul edilebilir güvenilirlik katsayısı sınırından yüksek olma özelliğini korurken, Müzik dersi için kabul edilebilir sınırların altında yer almaktadır. Öz değerlendirmeyle verilen puanların analize eklenmemesi durumunda ise; her üç ders kapsamında elde edilen G ve Φ katsayılarında artış görülmüştür. Bu artış diğer derslere göre Müzik dersinde daha fazla olmuştur. Akran değerlendirme puanlarının analize dahil edilmesinde; beş puanlayıcının verdikleri puanların ortalaması alınmıştır. Kısacası, beş puanlayıcı sanki bir puanlayıcı gibi davranmıştır. Bu durumda, akranlardan biri çok doğru puanlama yapamamış olsa bile diğerlerinin ortalamasıyla birlikte güvenirliliği yükseltmiş olabilir. Ama öz değerlendirmede öğrenciler sadece kendilerini puanladıkları için kendileri lehine puanlama yapmış da olabilirler. Öğrencilerin yaş özellikleri dikkate alındığında; arkadaşlarına daha yüksek puan vererek yanlış bir davranış sergilemek yerine olabildiğince öznel davranmaya özen göstermiş oldukları düşünülebilir. Bu durumda akran değerlendirmesi ve öğretmen değerlendirmesinin birbirine yakın çıkması, öz değerlendirmenin ise onlardan daha farklı bir konumda bulunması beklenebilir. Salmaner'in (2015) öz, öğretmen ve akran puanlarını çok yüzeyli rasch ölçme modeliyle incelediği araştırmasında da benzer bir sonucun ortaya çıktığı görülmektedir. Salmaner araştırmasında ilköğretim 5. sınıf öğrencileriyle çalışmış ve yaptığı analizlerin sonucunda en cömert puanlayıcıların öz puanlayıcılar olduğunu, en katı puanlayıcıların ise öğretmen veya akran puanlayıcılar olduğunu tespit etmiştir. Yaş grubu dikkate alındığında öğrencilerin başarılı olma isteklerinin veya başarısız olma kaygılarının kendilerine daha yüksek puan verme eğilimini yaratmış olabileceği söylenebilir.

Alanyazına bakıldığında; ilköğretim öğretmen, akran ve öz değerlendirmelerinin karşılaştırılmasına ilişkin yapılan çalışmalarda öğrencilerin henüz tam anlamıyla nesnel değerlendirmeler yapamadıkları; genellikle öz değerlendirmelerin en cömert puanlamalar olduğu görülmektedir (Salmaner, 2015; Sarıtaş, 2015). Börkan (2017) üniversite öğrencileriyle dört gün süren akran değerlendirme çalışmasında da dereceli puanlama anahtarı kullanarak öğrencilerin sunum performanslarını puanlamıştır. Çalışmanın sonucunda genel olarak akran puanlayıcıların arkadaşlarını oldukça cömert bir biçimde puanladıkları, puanlayıcılar kendi aralarında kıyaslandığında katılık/cömertlik seviyelerinin birbirlerinden farklı olduğu sonucuna ulaşılmıştır. Akran değerlendirme ve öz değerlendirmenin öğretmen değerlendirmesiyle birlikte ele alındığı bir başka çalışma da Matsuno (2009) tarafından yapılmıştır. Matsuno (2009) 19-21 yaşları arasında 91 Japon öğrencisi ve dört öğretmenle yaptığı bu çalışmada özellikle yüksek başarılı öğrenciler öz değerlendirme sürecinde kendi çalışmalarına tahmin edilenden daha düşük puan vermişler, akran değerlendirmesi sürecinde ise puanlayıcılar daha hoşgörülü ve tutarlı davranmışlardır. Kendi yazma yeteneklerinden bağımsız olarak yüksek başarılı öğrencileri düşük, düşük başarılı öğrencileri yüksek puanlamışlardır. Akran puanlayıcılarının çoğunun tutarlı olduğu, öz değerlendirme ve öğretmen puanlayıcılarından daha az önyargılı etkileşimler gösterdikleri görülmüştür. Farrokhi ve diğerleri (2011) ise çok yüzeyli Rasch modeli kullanarak öz değerlendirme, akran değerlendirme ve öğretmen değerlendirmesinde merkezîyetçilik eğilimini saptamak amacıyla bir çalışma yapmışlardır. 194 değerlendirici 188 adet yazılmış kompozisyonu altı analitik ölçekle değerlendirmişler ve araştırmanın sonunda akranlar arasında ve öz değerlendirme düzeyinde bir merkezîyetçilik olduğu sonucuna ulaşılmıştır. 2015 yılında Karakaya aday öğretmenlerin portfolyo dosyalarının değerlendirilmesinde öz değerlendirme, akran değerlendirme ve öğretmen değerlendirmeleri arasında bir karşılaştırma yapmıştır. Araştırmadan elde edilen bulgular öz değerlendirmede puanlayıcıların hoşgörülü davrandıklarını, akran değerlendirmede ise daha katı davrandıklarını belirlemiş; genel olarak değerlendiriciler arasında istatistiki olarak anlamlı bir fark bulmuştur. Nalbantoğlu-Yılmaz (2017) tarafından 56 öğretmen adayıyla yapılan bir diğer çalışmada ise yapılan bir projeye ilişkin öz değerlendirme ve akran değerlendirmelerinde farklılık olup olmadığının belirlenmesi ve öğretmen adaylarının kendi verdikleri notlar ile akranlarının ve öğretmenlerinin verdikleri puanların güvenilirliğinin ortaya çıkarılması amaçlanmıştır. Araştırmanın sonucunda değerlendirmeyi yapan kişiler arasında anlamlı bir fark bulunmamıştır. Öz değerlendirme, akran değerlendirme ve öğretmen değerlendirmelerinin güvenilirliğinin kabul edilebilir sınırlar içinde olduğunu ortaya koymuştur.

Araştırma sonuçlarından da görüldüğü gibi öğrencilerin kendi çalışmalarına ve akranlarının çalışmalarına yönelik değerlendirme yapmalarına daha çok fırsat verilerek onların eleştirel düşünme ve problem çözme gibi üst düzey düşünme becerilerini kullanmalarını sağlanmalıdır. Bu çalışma sonucunda da öğrencilerin öğrendiklerini günlük yaşamlarında kullanmalarına fırsat vermek, böylece öğrenilenleri içselleştirip kalıcı ve anlamlı öğrenmelerini sağlamak amacıyla örnek olay yöntemi gibi farklı bir öğretim yöntemi kullanılmış, öğrencilerin öğrenmelerini değerlendirmede alternatif değerlendirme yöntemlerinden öz değerlendirme ve akran değerlendirme becerilerini kullanmalarını sağlanarak, öğrenmeye ilişkin süreçleri derinlemesine anlamak için nasıl çaba gösterdikleri ortaya çıkarılmıştır.

Araştırmadan elde edilen bulgular öğrencilerin öz değerlendirme ve akran değerlendirme süreçlerini destekleyen; tartışma, eleştirel düşünme, problem çözme gibi üst düzey düşünme becerilerini geliştirilmesine yönelik etkinliklere daha fazla yer verilmesi gerektiğini göstermektedir. Bu becerilerin öğrencilere küçük yaşlardan itibaren kazandırılmasına ve bilimsel düşünebilen bireylerin yetiştirilmesine önem verilmelidir. Derslerde öz değerlendirme ve akran değerlendirme becerilerini geliştirmek amacıyla yapılan örnek olay yönteminin yanı sıra problem dayalı öğrenme, proje temelli öğrenme gibi farklı uygulamalara öğretim programlarında daha fazla yer verilmelidir. Disiplinler arası bağlantı kurularak öğretim programlarının daha etkili bir şekilde uygulanmasına katkı sağlanmalıdır.

Öğretim yöntemlerinin öğrencilerin düzeyine uygun seçilmesi, öğrencilerin öz değerlendirme ve akran değerlendirme becerilerini daha verimli kullanmalarını sağlayabilir. Öğrencilerin gelişim özellikleri dikkate alınarak, öğrenme konularına uygun değerlendirme ölçütleri öğrencilerle birlikte belirlenmelidir. Bunun için öğrencilerin alternatif değerlendirme yöntemlerine ilişkin daha fazla bilgi

sahibi olmaları sağlanmalıdır. Öğrencilere öz değerlendirme ve akran değerlendirmeye yönelik performans görevleri verilerek sorumluluk almalarına ve öğrenmelerine ilişkin farkındalık geliştirmelerine önem verilmelidir.

Bu çalışmada disiplinlerarası yaklaşıma dayalı olarak Türkçe, Hayat Bilgisi ve Müzik derslerinde puanlayıcı türleri arasındaki güvenilirlik incelenmiştir. Farklı araştırmalarda ders türleri ve sınıf düzeyi değiştirilebileceği gibi, tüm öğretmen değerlendirmelerinin aynı öğretmen tarafından yapılması sağlanabilir. Böyle bir çalışma ile elde edilen sonuçlarla dersler arasında yapılacak karşılaştırmalara karışacak hata kaynakları azaltılabilir. Çalışmada, çalışma grubunun büyüklüğü 30 olarak belirlenmiştir, ancak benzer çalışmalar daha büyük öğrenci grupları üzerinde tekrarlanabilir. Çalışmada Müzik dersi özelinde elde edilen güvenilirlik değerlerinin düşüklüğünün sebepleri farklı araştırmalarda detaylı olarak incelenebilir.

KAYNAKÇA

- Alicı, D. (2010). Öğrenci performansının değerlendirilmesinde kullanılan diğer ölçme araç ve yöntemleri. S. Tekindal (Ed.), *Eğitimde ölçme ve değerlendirme* (2. Baskı) içinde (ss.127-168). Ankara: Pegem Akademi.
- Atılğan, H. (2005). Genelenebilirlik kuramı ve puanlayıcılar arası güvenilirlik için örnek bir uygulama. *Eğitim Bilimleri ve Uygulama Dergisi*, 4(7), 95-108. http://www.ebuline.com/pdfs/7Sayi/7_6.pdf adresinden edinilmiştir.
- Bahar, M. (2006). *Fen ve teknoloji öğretimi*. Ankara: Pegem A Yayıncılık.
- Bahar, M., Nartgün, Z., Durmuş, S., & Bıçak, B. (2008). *Geleneksel-alternatif ölçme ve değerlendirme öğretmen el kitabı*. Ankara: Pegem A Yayıncılık.
- Ballantyne, R., Huges, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment and Evaluation in Higher Education*, 27(5), 427-441. doi: 10.1080/0260293022000009302
- Boud, D. (1986). *Implementing student self-assessment*. Sydney: Higher Education Research and Development Society of Australasia.
- Börkan, B. (2017). Akran değerlendirmesinde puanlayıcı katılığı kayması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(4), 469-489. doi: 10.21031/epod.328119
- Brennan, R. L. (2001). *Generalizability theory*. New York, NY: Springer-Verlag Inc.
- Büyükkıdık, S., & Anıl, D. (2015). Performansa dayalı durum belirlemede güvenilirliğin genelenebilirlik kuramında farklı desenlerle incelenmesi. *Eğitim ve Bilim*, 40(177), 285-296. doi: 10.15390/EB.2015.2454
- Cihanoğlu, M. O. (2008). *Alternatif değerlendirme yaklaşımlarından öz ve akran değerlendirmenin işbirlikli öğrenme ortamlarında akademik başarı, tutum ve kalıcılığa etkileri* (Yayımlanmamış doktora tezi). Dokuz Eylül Üniversitesi Eğitim Bilimleri Enstitüsü, İzmir.
- Cram, B. (1995). Self-assessment: From theory to practice. Developing a workshop guide for teachers. In G. Brindley (Ed.), *Language assessment in action* (pp. 271-350). Sydney: National Centre for English Language Teaching and Research, Macquarie University.
- Çeçen, M. A. (2011). Türkçe öğretmenlerinin seviye belirleme sınavı ve Türkçe sorularına ilişkin görüşleri. *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 8(15), 201-211. <http://www.acarindex.com/dosyalar/makale/acarindex-1423909379.pdf> adresinden edinilmiştir.
- Doğan, C. D., & Anadol, H. Ö. (2017). Genelenebilirlik kuramında tümüyle çaprazlanmış ve maddelerin puanlayıcılara yuvalandığı desenlerin karşılaştırılması. *Kastamonu Eğitim Dergisi*, 25(1), 361-372. <https://dergipark.org.tr/tr/pub/kefdergi/issue/27737/309180> adresinden edinilmiştir.
- Falchikov, N. (1986). Product comparisons and process benefits of collaborative peer group and self assessments. *Assesment and Evaluation in Higher Education*, 11(2), 146-166. doi: 10.1080/0260293860110206
- Falchikov, N. (2001). *Learning together; peer tutoring in higher education*. London: Routledge-Falmer.
- Farrokhi, F., Esfandiari R., & Dalili, M. V. (2011). Applying the many-facet rasch model to detect centrality in self-assessment, peer-assessment and teacher assessment. *World Applied Sciences Journal*, 15(Innovation and Pedagogy for Lifelong Learning), 70-77. Retrieved from <https://pdfs.semanticscholar.org/dd21/ba5683dde8b616374876b0c53da376c10ca9.pdf>
- Farrokhi, F., Esfandiari R., & Schaefer, E. (2012). A many-facet rasch measurement of differential rates severity/leniency in three types of assessment. *JALT Journal*, 34(1), 79-102. Retrieved from https://pdfs.semanticscholar.org/d79d/75e55050f9b977ffecd079ba5aadcdc10443.pdf?_ga=2.18401635.7.2134357192.1569916691-1527179006.1569916691

- Güler, N. (2009). Generalizability theory and comparison of the results of g and d studies computed by SPSS and Genova packet programs. *Education and Science*, 34(154), 93-103.
- Güler, N. (2011). Rasgele veriler üzerinde genellenabilirlik kuramı ve klasik test kuramı'na göre güvenilirliğin karşılaştırılması. *Education and Science*, 36(162), 225-234.
- Güler, N., Kaya-Uyanık, G., & Taşdelen Teker, G. (2012). *Genellenebilirlik kuramı*. Ankara: Pegem Akademi.
- İşman, A., & Eskicumalı, A. (2003). *Eğitimde planlama ve değerlendirme* (4. Baskı). İstanbul: Değişim Yayınları.
- Karakaya, İ. (2015). Comparison of self, peer and instructor assessments in the portfolio assessment by using many facet rasch model. *Journal of Education and Human Development*, 4(2), 182-192. doi: 10.15640/jehd.v4n2a22
- Kurudayıoğlu, M., Şahin, Ç., & Çelik, G. (2008). Türkiye'de uygulanan Türk edebiyatı programındaki ölçme ve değerlendirme boyutu uygulamasının değerlendirilmesi: Bir durum çalışması. *Ahi Evran Üniversitesi Kırşehir Eğitim Fakültesi Dergisi*, 9(2), 91-101. http://kefad.ahievran.edu.tr/InstitutionArchiveFiles/f44778c7-ad4a-e711-80ef-00224d68272d/d1a3a581-af4a-e711-80ef-00224d68272d/Cilt9Sayi2/JKEF_9_2_2008_91_101.pdf adresinden edinilmiştir.
- Kutlu, Ö., Doğan, C. D., & Karakaya, İ. (2008). *Öğrenci başarısının belirlenmesi, (performansa ve portfolyoya dayalı durum belirleme)*. Ankara: Pegem Akademi.
- Matsuno, S. (2009). Self-, peer-, and teacher-assessments in Japanese university EFL writing classrooms. *Language Testing*, 26(1), 75-100. doi: 10.1177/0265532208097337
- McMillan, H. J. (2015). *Sınıf içi değerlendirme*. (Çev. A. Arı). Ankara: Pegem A Yayıncılık.
- Millî Eğitim Bakanlığı. (2013). *İlköğretim Türkçe 3 öğretmen kılavuz kitabı*. Ankara: Millî Eğitim Bakanlığı.
- Millî Eğitim Bakanlığı. (2017a). *İlkokul hayat bilgisi öğretmen kılavuz kitabı 3. sınıf*. Ankara: Millî Eğitim Bakanlığı.
- Millî Eğitim Bakanlığı. (2017b). *İlköğretim müzik 4 öğretmen kılavuz kitabı*. Ankara: Millî Eğitim Bakanlığı.
- Mistar, J. (2011). A study of the validity and reliability of self-assessment. *Teflin Journal*, 22(1), 45-58. Retrieved from <http://journal.teflin.org/index.php/journal/article/viewFile/18/20>
- Nalbantoğlu-Yılmaz, F. (2017). Reliability of scores obtained from self-, peer-, and teacher-assessments on teaching materials prepared by teacher candidates. *Educational Sciences: Theory & Practice*, 17(2), 395-409. doi: 10.12738/estp.2017.2.0098
- Osterman, K. F., & Kottkamp, R. B. (1993). *Reflective practice for educators: Improving schooling through professional development*. Newbury Park, CA: Corwin Press.
- Race, P. (2001). *A briefing on self, peer and group assessment*, https://blogs.shu.ac.uk/teaching/files/2016/09/id9_briefing_on_self_peers_and_group_assessment_sna_s_901.pdf adresinden edinilmiştir.
- Salmaner, R. (2015). *Yazma becerilerinin değerlendirilmesinde öz akran ve öğretmen puanlarının çok yüzeyli rasch ölçme modeliyle incelenmesi* (Yayımlanmamış yüksek lisans tezi). Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Sarıtaş, S. (2015). *Problem çözme becerilerinin değerlendirilmesinde öz, akran ve öğretmen puanlarının çok yüzeyli Rasch ölçme modeli ile incelenmesi* (Yayımlanmamış yüksek lisans tezi). Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. USA: Sage Publications.
- Stiggins, J. R. (1997). *Student-centered classroom assessment*. New Jersey, NJ: Merrill Prentice Hall Inc.
- Stiggins, R., & Chappius, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory into Practice*, 44(1), 11-18. Retrieved from https://www.jstor.org/stable/3496986?seq=1#metadata_info_tab_contents
- Sünbül, A. M. (2007). *Öğretim ilke ve yöntemleri*. Konya: Çizgi Kitabevi.
- Taşdelen-Teker, G., Şahin, M. G., & Baytemir, K. (2016). Using generalizability theory to investigate the reliability of peer assessment. *Journal of Human Sciences*, 13(3), 5574-5586. Retrieved from <https://j-humansciences.com/ojs/index.php/IJHS/article/view/4155/2035>
- Taşdelen-Teker, G., & Güler, N. (2019). Thematic content analysis of studies using generalizability theory. *International Journal of Assessment Tools in Education*, 6(2), 279-299. doi: 10.21449/ijate.569996
- Tekindal, S. (2014). *Okullarda ölçme ve değerlendirme yöntemleri*. (4. Basım). Ankara: Nobel Akademik Yayıncılık.
- Topping, K. J., Smith, E. F., Swanson, I., & Elliot, A. (2000). Formative peer assessment of academic writing between postgraduate students. *Assessment and Evaluation in Higher Education*, 25(2), 149-169. doi: 10.1080/713611428
- Turgut, M. F., & Baykul, Y. (2015). *Eğitimde ölçme ve değerlendirme* (7. Baskı). Ankara: Pegem Akademi.

- Wilson, J., & Jan, W. L. (1993). *Thinking for themselves: Developing strategies for reflective learning*. Australia: Eleanor Curtain Publishing.
- Woolfolk, A. (2002). *Educational Psychology*. New York, NY: Pearson.
- Yaşar, M. (2017). Ölçme ve deđerlendirmenin önemi. S. Tekindal (Ed.), *Eđitimde ölçme ve deđerlendirme* (5. Baskı) içinde (ss.2-8). Ankara: Pegem Akademi.
- Yıldıztekin, B. (2014). *Klasik test kuramı ve genellebilirlik kuramından puanlayıcılar arası tutarlılıđın farklı yöntemlere göre karşılaştırılması* (Yayımlanmamış yüksek lisans tezi). Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara, Türkiye.