



İntramuskuler Enjeksiyon Yapma İstasyonu Verileriyle Genellenebilirlik Kuramında Dengelenmiş ve Dengelenmemiş Desenlerin Karşılaştırılması *

Funda Nalbantoğlu Yılmaz ¹, Ezel Tavşancıl ²

Öz

Araştırmanın genel amacı, İntramuskuler Enjeksiyon yapma istasyonu verileriyle, puanlanan öğrenci sayısı aynı kalmak şartıyla puanlayıcıların her birinin birbiriyle eşit ve birbirinden farklı sayıda öğrenciyi aynı görevler doğrultusunda puanlamasıyla oluşturulmuş dengelenmiş ve dengelenmemiş desenlerden elde edilen güvenilirlik ve varyans kestirimlerini karşılaştırmaktır. Araştırmanın çalışma grubunu, 2010-2011 öğretim yılı Hacettepe Üniversitesi Tıp Fakültesi Yapılandırılmış Objektif Klinik Sınav'daki İntramuskuler Enjeksiyon istasyonuna giren birinci sınıf öğrencilerinden 240 öğrenci oluşturmaktadır. Öğrencilerin ilgili istasyondaki performanslarının belirlenmesinde sekiz puanlayıcı kullanılmıştır. Puanlayıcıların her birinin birbiriyle eşit (dengelenmiş) ya da birbirinden farklı (dengelenmemiş) sayıda öğrenciyi puanlamasıyla oluşturulmuş desenlerle yapılan G çalışmaları sonucunda kestirilen varyans değerleri ve D çalışmasındaki senaryolara göre elde edilen G ve Phi katsayıları her iki desende de benzerlik göstermektedir.

Anahtar Kelimeler

Genellenebilirlik kuramı
Dengelenmiş desen
Dengelenmemiş desen

Makale Hakkında

Gönderim Tarihi: 06.11.2012
Kabul Tarihi: 23.10.2014
Elektronik Yayın Tarihi: 10.11.2014

DOI: 10.15390/EB.2014.2408

Giriş

Eğitimin her kademesinde ve personel seçiminde, zihin ve kas koordinasyonu gerektiren becerilerin değerlendirilmesinde performansın ölçülmesi oldukça önem taşımaktadır. Özellikle bir meslek grubuna ait becerilerin kazandırıldığı yükseköğretim programlarında ilgili beceriye ait performansın belirlenmesi daha da dikkati çekmektedir.

Bir duruma ait performans özünde bir görevi ya da süreci yürütme ve sonuçlandırmayı gerekli kılar (Wiggins, 1993). Kompozisyon yazma, bir müzik aleti çalma, sunum yapma gibi çeşitli faaliyetleri içerir. Böylece öğrenci bir süre boyunca bir performans durumuna ait bir ürün üretir ya da ilgili beceriyi yapar (Arias, 2010).

Tıp eğitiminde, öğrencilerin meslekleriyle ilgili klinik becerileri nasıl yaptıklarının değerlendirildiği performans sınavlarından biri de "Yapılandırılmış Objektif Klinik Sınav" (Objective Structure Clinical Examination) kısaca OSCE'dir. Bu sınav, birden fazla istasyondan oluşan ve her bir istasyonda farklı becerilerin puanlayıcılar tarafından değerlendirildiği bir performans sınavıdır (Elçin,

* Bu çalışma, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ölçme ve Değerlendirme Bölümü için hazırlanan doktora tezinden üretilmiştir.

¹ Nevşehir Hacı Bektaş Veli Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Türkiye, fundan@nevsehir.edu.tr

² Ankara Üniversitesi, Eğitim Bilimleri Fakültesi, Ölçme ve Değerlendirme Bölümü, Türkiye, etavsancil@gmail.com

Odabaşı ve Sayek, 2005). OSCE sınavında istasyonlar ise adayın farklı klinik becerilerini ölçmeye yönelik düzenlemiş farklı sınav ortamlarıdır.

OSCE sınavı tüm performans belirleme çalışmalarında olduğu gibi çeşitli faktörlerden etkilenmektedir. Bu faktörler sınavın yapıldığı ortamdan, süreden, ölçme aracından, puanlayıcının aşırı katı ya da cömert olmasından kaynaklı olabilmektedir. Değerlendirmeye karışabilecek bu tür hatalar nedeniyle elde edilen ölçme sonuçlarının güvenilirliği ve geçerliği, buna bağlı olarak da alınacak kararların tutarlılığı değişmektedir. Bu nedenle elde edilen performans puanlarını kullanmadan önce puanlara karışabilecek hata kaynaklarını dikkate alarak güvenilirlik incelemeleri yapmak büyük önem taşımaktadır.

Performansın ölçülmesinde güvenilirlik daha çok klasik test kuramı, madde tepki kuramı ve genellenebilirlik kuramına dayalı yöntemlerle çalışılmaktadır.

Klasik test kuramına göre güvenilirlik, gözlenen puanların gerçek puanları ne derece doğru yansıttığıdır. Klasik test kuramında güvenilirlik indeksi, gerçek puan varyansının gözlenen puan varyansına oranıdır. Bu güvenilirlik indeksi, bireylerin gerçek puanları bilinmediğinden pratik değildir (Webb, Shavelson ve Haertel, 2006). Bu nedenle klasik test kuramında güvenilirlik katsayısı çeşitli varsayımlara dayalı farklı yöntemler aracılığıyla dolaylı olarak elde edilmektedir.

Farklı kaynaklardan gelen hataların ölçme sonucuna karışabildiği durumlarda çeşitli güvenilirlik yöntemleriyle güvenilirliği değerlendirmek gerekir. Klasik test kuramında her hata kaynağına göre en az bir güvenilirlik tahmini vardır ve bu güvenilirlikler hata kaynağına bağlı olarak değişir. Klasik test kuramında tek bir gerçek puan ve ayırt edilmemiş bir hata terimi olmasına rağmen ölçme sonuçlarına karışan hata birden fazla kaynaktan gelebilmektedir (Cronbach, Gleser, Nanda ve Rajaratnam, 1972). Fakat klasik test kuramı bütün hata kaynaklarını tek değişkenlik kaynağından gelen hatalar olarak ele alır ve hata kaynaklarının birbirini etkilemediğini kabul eder. Klasik test kuramının aksine genellenebilirlik kuramı ise ölçme işlemindeki birçok hata kaynağını ve her bir hata kaynağının etkileşimlerini ele alan kavramsal çerçeve ve yöntem içermektedir (Brennan, 2001; Cronbach, Gleser, Nanda ve Rajaratnam, 1972; Shavelson ve Webb, 1991). Böylece genellenebilirlik kuramıyla performans belirleme çalışmalarında daha kapsamlı ve gerçekçi hata betimlemesi yapılabilir.

Genellenebilirlik kuramı, gözlenen puanlardaki hata kaynaklarının derecesinin belirlenmesini ve davranışın ölçülmesinde güvenilirliğin değerlendirilmesini, araştırılmasını sağlayan istatistiksel bir kuramdır (Brennan, 2001; Cronbach, Glaser, Nanda ve Rajaratnam, 1972; Shavelson ve Webb, 1991). Genellenebilirlik kuramının iki temel fonksiyonu vardır. İlki ölçme işlemlerinin niteliğini değerlendirmek, ikincisi ise ölçme işlemlerinin güvenilirliğinin nasıl artırılacağı hakkında kestirim yapmaktır (Wing ve Chiu, 2001). Bu çalışmaları yapabilmek için ise öncelikle çalışma desenleri oluşturulur. Oluşturulan bu desenler, araştırma verilerine göre kimi zaman dengelenmiş (balanced) kimi zaman da dengelenmemiş (unbalanced) olmaktadır.

Dengelenmiş desende gözlemlerin sayısı değişkenin her düzeyinde eşittir (Brennan, 2001). Örneğin; bir testin farklı okullardaki öğrencilere uygulandığı bir durumda öğrenciler okullar ile yuvalanmış ve her öğrencinin de tüm maddeleri aldığı (ö:o) x m deseninde, (ö: öğrenci, o:okul ve m:madde olmak üzere) her bir okuldan eşit sayıda öğrencinin uygulamaya katılması dengelenmiş bir durumdur. Üç alt testten oluşan bir testin tüm alt testlerinde aynı sayıda madde varsa bu desen yine dengeli bir desendir. Fakat pratikte dengelenmemiş desenlere ait durumlarla daha sık karşılaşılmaktadır. Çünkü yukarıda verilen örnekteki gibi, her bir okuldan eşit sayıda öğrencinin uygulamaya alınması ya da alt testlerden oluşan bir testin her bir alt testindeki madde sayısının aynı olması her durumda beklenemez. Dengelenmiş desenlerde, değişkenin her bir düzeyi için gözlemlerin sayısının aynı sayıda olduğu belirtilmişti. Fakat her durumda değişken için elde edilen gözlemler eşit sayıda olmayabilir. Kayıp (missing) veri ya da bir değişkenin düzeylerine ait gözlem sayısının değiştiği durumlarda dengelenmemiş desenler ortaya çıkmaktadır (Brennan, 2001).

Bir değişkenin koşullarına ait hücrelerdeki gözlem sayısının eşit olmadığı durumlarda veri dengelenmemiş olarak adlandırılır (Kaufmann ve Schering, 2007). Değişkenin tüm koşullarına ait elde edilen gözlemlerin eşit sayıda olmadığı durumlardaki verilerden dengelenmiş veri yaratmak için bazı verilerin atılması gerekir. Bu pratik bir yöntem değildir. Çünkü bazı durumlarda verinin büyük miktarının atılması gerekebilir. Bu da çok fazla veri kaybına neden olur. Böyle durumlarda veri atmaya yerine dengelenmemiş desenlerle varyans bileşenlerini hesaplamak gerekir.

Literatürdeki genellenebilirlik kuramıyla güvenirliliğin araştırıldığı çalışmalar incelendiğinde, yurt dışında yapılanların büyük çoğunluğunda dengelenmiş veri yapılarının kullanılmasının yanında, dengelenmemiş verilerle ve her iki veri yapısının bir arada kullanıldığı çalışmaların olduğu görülmektedir (Jeon, Lee, Hwang ve Kang, 2009; Lee ve Frisbie 1999; Ødegård, Hagtvet ve Bjørkly, 2008; Sharma ve Weathers, 2003; Wei and Haertel, 2011). Türkiye’de genellenebilirlik kuramının kullanıldığı araştırmalar incelendiğinde ise büyük çoğunluğunun genellenebilirlik kuramının klasik test kuramı ve/veya Rasch modeli ile karşılaştırılmasına yönelik ağırlık kazandığı görülmektedir (Atılğan, 2008; Deliceoğlu ve Çıkrıkçı Demirtaşlı, 2012; Güler ve Gelbal, 2010; Güler, 2011; Taşdelen, 2009; Yelboğa ve Tavşancıl, 2010; Yılmaz Nalbantoğlu ve Gelbal, 2011). Ayrıca yapılan bu çalışmaların tümünde dengelenmiş desenlerin kullanıldığı da dikkati çekmektedir. Bununla birlikte dengelenmemiş desenlerin çoğunun gerçek durum verilerine daha uygun olduğu bilinmektedir. Uygulama şartlarında, araştırmacıların sahip oldukları verilerin her zaman dengeli yapıda olması beklenemez. Özellikle OSCE sınavı gibi öğrencilerin birden fazla performans durumlarının ölçüldüğü ve/veya öğrenci sayısının fazla olduğu sınavlarda her bir öğrencinin performansını birden fazla puanlayıcı kullanarak değerlendirmek, zaman ve iş gücü açısından bazı sınırlılıklar taşımaktadır. Bu sınırlılıklar nedeniyle OSCE sınavında her bir istasyonda tek puanlayıcı kullanılmakta ve puanlayıcılar belli aralıklarla değişmektedir. Puanlayıcıların öğrencileri dönüşümlü puanlaması nedeniyle de her bir puanlayıcının puanladığı toplam öğrenci sayısı farklılaşmakta, puanlayıcıların puanladığı öğrenci sayısı bakımından dengelenmemiş veri yapısı oluşmaktadır. Bu doğrultuda, bu araştırma ile mevcut uygulamaya ait güvenirliliğin belirlenmesinin, puanlayıcıların her birinin birbirleriyle eşit ve birbirlerinden farklı sayıda öğrenciyi puanlamasının güvenirliliğe olan etkisinin incelenmesinin OSCE sınavı uygulamalarına katkı sağlayacağı düşünülmektedir. Ayrıca araştırmanın çok sayıda öğrencinin ve/veya birden fazla puanlayıcının yer aldığı ve performansa dayalı kararların verildiği güzel sanatlar, beden eğitimi vb. farklı alanlardaki diğer araştırmacılara da farklı kaynaklardan gelen hataları ve bu hataların etkileşimlerini de dikkate alarak güvenirlilik belirleme ve puanlayıcı etkisinin araştırılması bakımından da katkı sağlayacağı düşünülmektedir. Böylece Türkiye’de genellenebilirlik kuramıyla performans belirleme çalışmalarında güvenirliliğin araştırıldığı çalışmaların sınırlı olması ve pratikte daha çok karşılaşılan bir durum olan dengelenmemiş veri yapılarındaki desenlerle örnek uygulamanın olmaması bu karşılaştırmayı ortaya koyan çalışmaların yapılması için bir gereklilik olarak görülmüştür.

Çalışmada dengelenmiş ve dengelenmemiş veri yapılarıyla oluşturulan desenlerden elde edilen güvenirliliğin nasıl olacağı, aynı verilerle dengelenmemiş desenlerle hesaplanan güvenirlilik katsayılarının dengeli yapılara göre güvenirliliği etkileyip etkilemeyeceği ve desenler için kestirilen varyans bileşenleri arasında farklılık olup olmayacağı belirlenmesi amaçlanmıştır. Bu doğrultuda araştırmanın genel amacı, tıp eğitiminde öğrencilerin mesleki becerilerinin değerlendirildiği OSCE sınavı İntramusküler Enjeksiyon yapma istasyonu verileriyle, puanlanan öğrenci sayısı aynı kalmak şartıyla puanlayıcıların her birinin birbiriyle eşit ve birbirinden farklı sayıda öğrenciyi aynı görevler doğrultusunda puanlamasıyla oluşturulmuş dengelenmiş ve dengelenmemiş desenlerle yapılan genellenebilirlik kuramı analizlerinden elde edilen sonuçları karşılaştırmaktır. Bu amaç doğrultusunda ise aşağıdaki sorulara cevap aranmıştır.

1. Dengelenmiş ve dengelenmemiş (ö:p) x g desenlerinden genellenebilirlik kuramıyla elde edilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nedir?
2. Her iki desende puanlayıcı ve öğrenci sayılarının artırılıp azaltılmasıyla oluşturulan senaryolara göre yapılan karar çalışmalarından kestirilen G ve Phi katsayıları nasıl değişmektedir?

Yöntem

Araştırmanın Türü

Bu çalışmada, puanlayıcıların dönüşümlü olarak öğrencilerin performanslarını değerlendirdiği klinik sınav verileriyle oluşturulmuş dengelenmiş ve dengelenmemiş desenlerin nasıl sonuçlar verdiği incelenmiş ve her iki desenden elde edilen sonuçlar karşılaştırılmıştır. Bu yönüyle araştırma temel araştırma niteliğindedir.

Çalışma Grubu

Araştırmanın çalışma grubunu, 2010-2011 öğretim yılı Hacettepe Üniversitesi Tıp Fakültesi Yapılandırılmış Objektif Klinik Sınav'daki İntramusküler Enjeksiyon istasyonuna giren ve değerlendirme formundaki bilgilerin tam doldurulduğu birinci sınıf öğrencilerinden seçilen 240 öğrenci oluşturmaktadır. Öğrencilerin ilgili istasyondaki performanslarının değerlendirilmesinde Yapılandırılmış Objektif Klinik sınav kapsamında 12 puanlayıcı kullanılmaktadır. Puanlayıcıların her birinin puanladığı öğrenci sayısı birbirlerine göre farklılık göstermektedir. İlgili istasyonuna ait veri puanlayıcıların birbirlerinden farklı sayıda öğrenciyi puanlaması bakımından dengelenmemiştir. Çalışma kapsamında dengelenmiş ve dengelenmemiş desenlerden elde edilen sonuçların karşılaştırılması amaçlanmaktadır. Bu nedenle her iki desende puanlanan toplam öğrenci sayısı eşit olacak şekilde her bir puanlayıcının puanladığı öğrenci sayısını dikkate alarak dengelenmemiş ve dengelenmiş veri yapıları oluşturabilmek için çalışma kapsamında sekiz puanlayıcının puanlamalarından yararlanılmıştır. Öğrencilerin İntramusküler Enjeksiyon yapma becerilerinin puanlanmasında görev alan bu puanlayıcılar Tıp Eğitimi ve Bilişimi Anabilim Dalı tarafından belirlenmiş olup puanlayıcıların hepsi ilgili alandan seçilmiştir.

Araştırma Verileri

Araştırmada, Hacettepe Üniversitesi Tıp Fakültesi, Tıp Eğitimi ve Bilişimi Anabilim Dalı'nın 2010-2011 akademik yılına ait birinci dönem öğrencilerinin Yapılandırılmış Objektif Klinik Sınavındaki İntramusküler Enjeksiyon yapma istasyonundan elde edilen veri kullanılmıştır.

İntramusküler Enjeksiyon istasyonunda öğrenciler sırayla maket üzerinde İntramusküler Enjeksiyon yapmaktadır. İstasyonda her bir öğrenciye eşit süre verilmektedir. Öğrencilerin ilgili istasyondaki becerilerini değerlendirmek için istasyonda bir puanlayıcı bulunmaktadır. Bu puanlayıcı kendisine ayrılan süre içinde sırayla istasyona giren öğrencileri gösterdikleri becerilere göre Tıp Eğitimi ve Bilişimi Anabilim Dalı tarafından bu istasyonda kullanılmak üzere hazırlanan ve 17 görevden oluşan İntramusküler Enjeksiyon Yapma Becerisi değerlendirme formunu kullanarak değerlendirmektedir. İstasyonda dönüşümlü olarak bulunan her bir puanlayıcı kendisi için ayrılan süre bittiğinde istasyondan çıkmakta ve yerine başka bir puanlayıcı istasyona girmektedir. Böylelikle belli aralıklarla istasyondaki puanlayıcılar yer değiştirmekte ve her bir puanlayıcı birbirlerinden farklı bir grup ve farklı sayıda öğrenciyi puanlamaktadır.

Elde edilen araştırma verileriyle, genellenebilirlik kuramı analizleri için her bir puanlayıcının dönüşümlü olarak araştırmaya katılan öğrencilerden sadece bir kısmını ve birbirlerinden farklı bir grup öğrenciyi puanlamasıyla, öğrenci (ö) ve puanlayıcı (p) değişkenlerinin yuvalanmış, tüm öğrenciler için aynı olan görevlerin (g) ise bu değişkenlerle çaprazlandığı (ö:p) x g deseni dengelenmiş ve dengelenmemiş durumlara göre incelenmiştir. Araştırma kapsamında incelenen dengelenmiş ve dengelenmemiş (ö:p) x g desenleri arasındaki fark, puanlayıcıların her birinin puanladığı öğrenci sayılarının dengelenmiş desende birbirlerine eşit, dengelenmemiş desende ise birbirlerinden farklı olmasından kaynaklanmaktadır. Dengelenmiş (ö:p) x g deseninde sekiz puanlayıcıdan her biri birbirleriyle eşit sayıda olmak üzere 30'ar öğrenciyi 17 görev doğrultusunda puanlamaktadır. Dengelenmemiş (ö:p) x g deseninde ise 240 öğrenciden 23 öğrenci 1. puanlayıcı, 25 öğrenci 2. puanlayıcı, 26 öğrenci 3. puanlayıcı, 28 öğrenci 4. puanlayıcı, 30 öğrenci 5. puanlayıcı, 31 öğrenci 6. puanlayıcı, 36 öğrenci 7. puanlayıcı ve 41 öğrenci de 8. puanlayıcı tarafından 17 görev doğrultusunda puanlanmaktadır.

Verilerin Analizi

Araştırma verilerinin analizinde, dengelenmiş ve dengelenmemiş verilerin genellenebilirlik kuramı analizlerinde kullanılan urGENOVA (Brennan, 2001) programı temelli G_String (G-string-IV, Version 6.1.1.) ara yüz programı kullanılmıştır.

Bulgular

Bu bölümde, araştırmanın genel amacı çerçevesinde alt amaçlara göre elde edilen bulgulara yer verilmiştir.

Dengelenmiş ve Dengelenmemiş Desenlerden Elde Edilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Aynı veriler kullanılarak puanlayıcıların her birinin birbirlerine eşit ve birbirlerinden farklı sayıda öğrenciyi dönüşümlü olarak puanlamasına ait iki durumda (ö:p) x g desenine göre kestirilmiş varyans ve toplam varyansı açıklama yüzdeleri Tablo 1’de gösterilmiştir.

Tablo 1. Dengelenmiş ve Dengelenmemiş (ö:p) x g Desenlerinden Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Varyans Bileşeni	Sd	Dengelenmiş (ö:p) x g Deseni		Dengelenmemiş (ö:p) x g Deseni	
			σ^2	%	σ^2	%
g	σ^2_g	16	0.00141	2.6	0.00142	2.5
p	σ^2_p	7	0.00043	0.8	0.00032	0.6
ö : p	$\sigma^2_{ö:p}$	232	0.00199	3.6	0.00193	3.4
gp	σ^2_{gp}	112	0.00514	9.4	0.00541	9.6
gö : p	$\sigma^2_{gö:p,e}$	3712	0.04571	83.6	0.04712	83.8
Toplam		4079		100		100

ö: öğrenci, p: puanlayıcı, g: görev

Tablo 1’deki verilere göre, görevlere ait varyans bileşenlerinin toplam varyansı açıklama yüzdesi, dengelenmiş (ö:p) x g deseninde %2.6 ve dengelenmemiş (ö:p) x g deseninde %2.5 olarak hesaplanmıştır. Her iki desenle görevlere ait kestirilen varyans bileşenlerinin düşük olduğu görülmektedir. Bu nedenle her iki desen için de görevlerin zorluk-kolaylık bakımından farklılaşmadığı, görevlerin birbirlerine göre öğrencilere eşit güçlükte geldiği yorumu yapılabilir.

Puanlayıcı ana etkisi için kestirilen varyans bileşeni dengelenmiş (ö:p) x g deseninde 0.00043, dengelenmemiş (ö:p) x g deseninde ise 0.00032 olarak kestirilmiştir. Ayrıca her iki desenle kestirilen puanlayıcı etkisine ait varyans bileşeninin toplam varyansı açıklama yüzdesi dengelenmiş desende %0.8, dengelenmemiş desende ise %0.6’dır. Kestirilen bu varyans değerlerinin ve toplam varyansı açıklama yüzdelerinin düşük ve birbirine yakın olduğu görülmektedir. Bu bağlamda, her iki desen için puanlayıcı değişkenliğinin öğrencilerin performansını etkilemediği, puanlayıcıların puanlama bakımından benzer olduğu yorumu yapılabilir.

Puanlayıcıların aynı görevler doğrultusunda öğrencileri dönüşümlü olarak puanlamasıyla oluşturulmuş dengelenmiş ve dengelenmemiş (ö:p) x g desenlerinde öğrencilerin puanlayıcılarla yuvalandığı (ö:p) değişkene ait varyans bileşeni, dengelenmiş desende 0.00199 (%3.6), dengelenmemiş desende ise 0.00193 (%3.4) olarak kestirilmiştir. Elde edilen sonuçlardan da görüldüğü gibi ö:p değişkenine ait varyans kestirimleri ve bu kestirimlerin toplam varyansı açıklama yüzdeleri her iki desende de birbirine yakın ve düşük çıkmıştır. Bu durum her iki desen için de öğrencilerin İntramusküler Enjeksiyon yapma becerileri bakımından farklılaşmadığı ve puanlayıcıların her birinin davranışlarının bir öğrenciden diğerine değişmediğini göstermektedir.

Dengelenmiş ve dengelenmemiş (ö:p) x g desenlerinde görev x puanlayıcı (gxp) ortak etkileşimine ait varyans bileşenleri dengelenmiş desende 0.00514 (%9.4), dengelenmemiş desende ise 0.00541 (%9.6) olarak kestirilmiştir. Elde edilen bu bulgulardan her iki desenle kestirilen varyans bileşenlerinin birbirine çok yakın olduğu ve her iki desende de görev x puanlayıcı ortak etkisinin artık

varyans hariç diğer varyans bileşenlerine oranla yüksek olduğu görülmektedir. Bu nedenle, her iki desen de görev x puanlayıcı ortak etkisinden kaynaklı farklılıkların olduğu söylenebilir.

Tablo 1 incelendiğinde puanlayıcıların birbirleriyle eşit sayıda öğrenciyi puanladığı dengelenmiş (ö:p) x g deseninde artık varyansın ($\sigma^2(\text{gö:p})$) toplam varyansı açıklama yüzdesinin %83.6 olduğu, puanlayıcıların birbirinden farklı sayıda öğrenciyi puanladığı dengelenmemiş (ö:p) x g deseninde ise artık varyansın ($\sigma^2(\text{gö:p})$) toplam varyansın %83.8'ini açıkladığı görülmektedir. Her iki desende de artık varyans yüksek çıkmıştır. (ö:p) x g desenlerinde artık varyansın yüksek olması öğrenci x görev ortak etkileşimi, öğrenci x görev x puanlayıcı etkileşiminden gelen farkların ve/veya diğer açıklanamayan değişkenlik kaynaklarından gelebilecek etkilerin yüksek olabileceğinin bir göstergesidir. Bu nedenle her iki desen için öğrenci x görev ortak etkileşimi, öğrenci x görev x puanlayıcı etkileşimi ve/veya açıklanamayan değişkenlik kaynaklarından gelen etkilerin olduğu söylenebilir.

Her İki Desende Puanlayıcı ve Öğrenci Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışmalarının Karşılaştırılması

Araştırma amacı doğrultusunda karşılaştırılan dengelenmiş ve dengelenmemiş durumlar puanlayıcıların puanladığı öğrenci grubundaki öğrenci sayılarının birbirlerine eşit ve birbirlerinden farklı olmasından kaynaklanmaktadır. Bu açıdan yapılan karar çalışmasında, öğrencilerin farklı puanlayıcılar tarafından aynı görevler doğrultusunda dönüşümlü olarak puanlanmasıyla görevlerin ölçmenin objesi ve tüm değişkenlerin tesadüfi olarak belirlendiği (ö:p) x g deseninde evren puanı görevler olarak belirlenmiş, puanlayıcı ve öğrenciler ise genelleme evreni olarak alınmıştır. Bu doğrultuda puanlayıcıların her birinin birbirlerine eşit (dengelenmiş) ve birbirlerinden farklı (dengelenmemiş) sayıda öğrenciyi puanlamasına ilişkin desenlerde, öğrenci ve puanlayıcı sayılarının artırılıp azaltılmasıyla yapılan karar çalışmalarından elde edilen G ve Phi katsayıları Tablo 2'de karşılaştırılmıştır.

Tablo 2'de verilen senaryolara göre dengelenmemiş desende 240 öğrencinin sekiz puanlayıcı tarafından ve her bir puanlayıcının puanladığı öğrenci sayıları arasındaki farklılığın uygulamadaki dengelenmemiş veriye göre azaltıldığı durumda ($n_{\text{ö}+}:240, n_p:8, \check{n}_p:7.96$) G katsayısı 0.618, Phi katsayısı ise 0.605 olarak kestirilmiştir. Elde edilen bu değerler, uygulamadaki dengelenmiş ve dengelenmemiş desenden elde edilen katsayılarla karşılaştırıldığında dengelenmiş desenden ($n_{\text{ö}+}:240, n_p:8$) daha düşük, dengelenmemiş desenden ($n_{\text{ö}+}:240, n_p:8, \check{n}_p:7.73$) ise daha yüksek olduğu görülmektedir. Bu durum, puanlayıcıların puanladığı öğrenci sayısı bakımından, puanlayıcılar arası değişkenliğin azaltıldığında G ve Phi katsayılarının yükselme eğiliminde olduğunu göstermektedir.

Tablo 2. Dengelenmiş ve Dengelenmemiş (ö:p) x g Desenleri İle Senaryolara Göre Yapılan Karar Çalışmalarından Elde Edilen G ve Phi Katsayılarının Karşılaştırılması

Dengelenmiş (ö:p) x g Deseni					Dengelenmemiş (ö:p) x g Deseni					
$n_{\text{ö}+}$	n_p	$n_{\text{ö:p}}$	G	Phi	$n_{\text{ö}+}$	n_p	$n_{\text{ö:p}}$	\check{n}_p	G	Phi
					120	6	16,18,20,21,22,23	5.92	0.520	0.507
120	6	20	0.533	0.515	120	8	10,12,14,15,15,17,18,19	7.73	0.562	0.549
120	8	15	0.579	0.563	120	8	13,13,15,15,15,15,17,17	7.93	0.568	0.556
120	10	12	0.612	0.596	120	10	8,9,10,11,12,12,12,14,15,17	9.55	0.592	0.580
240	6	40	0.574	0.556	240	6	33,38,40,41,43,45	5.95	0.562	0.548
240	8	30	0.629	0.612	240	8	23,25,26,28,30,31,36,41	7.73	0.613	0.600
240	10	24	0.667	0.651	240	8	26,28,30,30,30,30,33,33	7.96	0.618	0.605
360	6	60	0.589	0.571	240	10	15,17,20,22,24,26,27,28,30,31	9.56	0.648	0.636
360	8	45	0.647	0.630	360	6	55,58,60,60,62,65	5.98	0.578	0.565
360	10	36	0.687	0.672	360	8	38,40,43,45,47,48,49,50	7.94	0.636	0.623
					360	8	41,43,45,45,45,48,48	7.98	0.637	0.624
					360	10	22,26,31,31,35,36,38,39,45,57	9.36	0.664	0.652

$n_{\text{ö}+}$: toplam öğrenci sayısı n_p : puanlayıcı sayısı, $n_{\text{ö:p}}$: her bir puanlayıcının puanladığı öğrenci sayısı, \check{n}_p : $n_{\text{ö}+}/\sum n_{\text{ö:p}}$

Her iki desene göre verilen senaryo durumlarında bu kez uygulamadaki toplam öğrenci sayısı sabit tutularak ($n_{\text{ö}}=240$) puanlayıcı sayısı azaltıldığında ($n_{\text{p}}=6$) dengelenmiş desende G katsayısı 0.574, Phi katsayısı 0.556 olarak kestirilmiştir. Aynı durumda dengelenmemiş desende ise G katsayısı 0.562, Phi katsayısı 0.548 olarak kestirilmiştir. Her iki desen için yine puanlanan toplam öğrenci sayısı sabit kalmak koşuluyla puanlayıcı sayısı artırıldığında ($n_{\text{p}}=10$) dengelenmiş desende G katsayısı 0.667, Phi katsayısı 0.651, dengelenmemiş desende ise G katsayısı 0.648, Phi katsayısı da 0.636 olarak kestirilmiştir. Bu bulgulardan hareketle, puanlayıcı sayısının azaltılıp artırılması ile elde edilen artış miktarının uygulamadaki ($n_{\text{ö}}=240$, $n_{\text{p}}=8$) verilere göre az olduğu görülmektedir. Ayrıca toplam öğrenci sayısı sabit tutularak puanlayıcı sayısının azaltılması ve artırılması durumlarında puanlayıcıların her birinin birbiriyle eşit ve birbirinden farklı sayıda öğrenciyi puanladığı iki desenden elde edilen katsayılar arasındaki farkın da düşük olduğu söylenebilir.

Tablo 2'deki verilerden bu kez puanlayıcı sayısı sabit tutularak ($n_{\text{p}}=8$) öğrenci sayısı azaltıldığında ($n_{\text{ö}}=120$) dengelenmiş desende G katsayısı 0.579, Phi katsayısı ise 0.563 olarak kestirilmiştir. Aynı durumda dengelenmemiş desende ($n_{\text{ö}}=120$, $n_{\text{p}}=8$, $\check{n}_{\text{p}}=7.73$) G katsayısı 0.562, Phi katsayısı da 0.549 olarak kestirilirken; puanlayıcıların puanladığı öğrenci sayısı bakımından dengesizlik azaltıldığında ($n_{\text{ö}}=120$, $n_{\text{p}}=8$, $\check{n}_{\text{p}}=7.93$) ise G katsayısı 0.568, Phi katsayısı da 0.556 olarak kestirilmiştir. Puanlayıcı sayısı uygulamadaki gibi sabit tutulup öğrenci sayısı artırıldığında ($n_{\text{ö}}=360$) ise dengelenmiş desene göre G katsayısı 0.647, Phi katsayısı 0.630 olarak kestirilirken; dengelenmemiş desene göre ($n_{\text{ö}}=360$, $n_{\text{p}}=8$, $\check{n}_{\text{p}}=7.94$) ise G katsayısı 0.636, Phi katsayısı da 0.623 olarak kestirilmiştir. Ayrıca puanlayıcıların puanladığı öğrenci sayısı bakımından dengesizlik azaltıldığında ($n_{\text{ö}}=360$, $n_{\text{p}}=8$, $\check{n}_{\text{p}}=7.98$), G katsayısı 0.637, Phi katsayısı da 0.624 olarak kestirilmiştir. Böylece, puanlayıcı sayısı sabit tutularak puanlanan toplam öğrenci sayısının azaltılması ve artırılması durumlarında elde edilen katsayıların her iki desende birbirine yakın olduğu görülmektedir.

Her iki desende puanlayıcı ve öğrenci sayısının uygulamadaki verilere göre birlikte artırılması durumunda ($n_{\text{ö}}=360$, $n_{\text{p}}=10$), dengelenmiş (ö:p) x g deseninde G katsayısı 0.687, Phi katsayısı 0.672, dengelenmemiş (ö:p) x g deseninde ise G katsayısı 0.664, Phi katsayısı 0.652 olarak kestirilmiştir. Elde edilen bu bulgulardan hareketle, öğrenci ve puanlayıcı sayılarının birlikte artırılması durumunda uygulamadaki verilere göre G ve Phi katsayılarının bir miktar arttığı ve her iki desendeki katsayıların birbirine yakın sonuçlar ürettiği yorumu yapılabilir.

Her iki desene ait karar çalışmasında, sekiz puanlayıcının birbirlerinden farklı sayıda olmak üzere toplamda 240 öğrenciyi 17 görev doğrultusunda puanlamasıyla dengelenmemiş desenden elde edilen G ve Phi katsayıları, altı puanlayıcının toplamda 120 öğrenciyi ilk durumdakiyle aynı görevler doğrultusunda puanlamasıyla dengelenmiş desenden elde edilen G ve Phi katsayılarından daha yüksek çıkmıştır. Aynı durum dengelenmemiş desende 360 öğrencinin sekiz puanlayıcı tarafından puanlanması ile dengelenmiş desende 120 öğrencinin sekiz puanlayıcı tarafından puanlanmasıyla elde edilen katsayılar arasında da görülmektedir. Bu durum dengelenmemiş veri yapılarından veri indirgemesi yapılarak dengelenmiş desenlerle güvenilirlik katsayıları hesaplandığında her iki desende ki katsayıların farklılaştığını ve çok fazla veri kaybı olduğunda güvenilirliğin düştüğünü göstermektedir.

Tartışma, Sonuç ve Öneriler

OSCE sınavı İnamusküler Enjeksiyon yapma istasyonu için G çalışmasıyla dengelenmiş ve dengelenmemiş (ö:p) x g desenlerinden elde edilen varyans ve toplam varyansı açıklama yüzdelerine ait sonuçlar aşağıda verilmiştir.

İnamusküler Enjeksiyon yapma istasyonu için puanlayıcıların her birinin birbiriyle eşit ya da birbirinden farklı sayıda öğrenciyi puanlamasıyla oluşturulmuş dengelenmiş ve dengelenmemiş (ö:p) x g desenleri ile yapılan G çalışmaları sonunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri incelendiğinde, her iki desenle kestirilen varyans bileşenlerinin ve toplam varyansı açıklama yüzdelerinin benzer sonuçlar ürettiği tespit edilmiştir. Araştırmadan elde edilen bu bulgu, Shavelson ve Webb'in (1981), mesleklere ait açıklamaların farklı coğrafi bölgeden gelen puanlayıcılar tarafından değerlendirildiği bir durumda, her bir coğrafi bölgeden gelen puanlayıcıların eşit ya da farklı sayıda olmasına göre oluşturulan dengelenmiş ve dengelenmemiş desenlerden elde edilen varyans kestirimlerinin benzerlik gösterdiği sonuçlarıyla paralellik göstermektedir. Ayrıca Sharma ve Weathers'ın (2003), her bir ülkeden eşit sayıda bireyin alındığı dengelenmiş ve her bir ülkeden farklı sayıda bireyin alındığı dengelenmemiş verilerle varyans bileşenlerine ait kestirimlerin hesaplandığı çalışmalarında da her iki durumdaki varyans kestirimlerinin benzer olduğu tespit edilmiştir.

Araştırmada kullanılan desenlerden elde edilen varyans kestirimleri dikkate alındığında ise, İnamusküler Enjeksiyon yapma istasyonuna ait görevlerin zorluk-kolaylık bakımından farklılaşmadığı, görevlerin öğrencilere eşit güçlükte geldiği, puanlayıcı değişkenliğinin öğrencilerin performansını etkilemediği, puanlayıcıların puanlama bakımından farklılığa neden olmadığı, puanlayıcıların her birinin davranışlarının bir öğrenciden diğerine değişmediği sonuçlarına ulaşılmıştır. Puanlayıcı değişkenliğinin öğrencilerin performansını etkilemediği, puanlayıcıların puanlama bakımından farklılığına neden olmadığı bulgusu ise Yılmaz Nalbantoğlu ve Gelbal (2011), Yılmaz Nalbantoğlu ve Başusta Uzun (2012) tarafından Hacettepe Üniversitesi OSCE sınavına ait farklı becerilerin değerlendirildiği farklı istasyonlara ait çalışmaların bulgularıyla benzerlik göstermektedir.

OSCE sınavı İnamusküler Enjeksiyon yapma istasyonu için karar çalışmasında dengelenmiş ve dengelenmemiş (ö:p) x g desenleri ile elde edilen G ve Phi katsayılarına ait sonuçlar ise aşağıda verilmiştir.

Yapılan karar çalışmaları sonucunda her iki desenle elde edilen G ve Phi katsayılarının yüksek olmadığı görülmüştür. Bunun nedeni olarak görevlerin zorluk ve kolaylık bakımından farklılaşmaması, öğrencilerin görevlerdeki performanslarının farklılık göstermemesi, grubun ölçülen özellik bakımından homojen olması gösterilebilir.

Puanlayıcıların birbirlerine eşit ya da birbirlerinden farklı sayıda öğrenciyi aynı görevler doğrultusunda puanlamasıyla, toplam veri sayısı dengelenmiş ve dengelenmemiş durumlarda aynı kalmak şartıyla uygulamadaki verilere ve çeşitli senaryo durumlarına göre hesaplanan katsayılar her iki desende birbirine yakın sonuçlar üretmektedir. Araştırmadan elde edilen bu bulgu Jeon, Lee, Hwang ve Kang (2009), Lee ve Frisbie (1999), Malhotra ve Sharma (2008), Shavelson ve Webb (1981) tarafından dengelenmiş ve dengelenmemiş desenlerle farklı çalışmalarla elde edilen güvenilirlik katsayılarının her iki desende birbirlerine yakın sonuçlar verdiği bulgularıyla da tutarlılık göstermektedir. Fakat araştırmada yapılan karar çalışması, kullanılan veri yapısının doğası gereği dengelenmemiş yapıda olması ve bu nedenle veri indirgemesi yapılarak veri kaybıyla dengelenmemiş durumdan dengelenmiş veri yapısı elde edildiğinde her iki desenden elde edilen güvenilirlik katsayılarının farklılaştığını, dengelenmemiş durumda daha yüksek güvenilirlik katsayısı elde edildiğini göstermektedir. Bununla birlikte puanlanan toplam öğrenci sayısı sabitken puanlayıcıların her birinin puanladığı öğrenci sayısı farklılaştıkça, bir başka deyişle puanlayıcıların puanladığı öğrenci sayıları arasındaki değişkenlik arttıkça güvenilirlik katsayılarının düşme eğiliminde olduğu da tespit edilmiştir.

Sonuç olarak, psikomotor becerilerin ağırlıkta olduğu bu çalışmada puanlayıcıların her birinin birbiriyle eşit (dengelenmiş) ve birbirinden farklı sayıda (dengelenmemiş) öğrenciyi dönüşümlü olarak puanlanmasına ait iki durumda toplam veri sayısı iki durumda da aynı olduğunda elde edilen güvenilirlik katsayıları birbirine yakın olmakla birlikte veri indirgemesi yapıldığında ya da dengesizlik arttığında güvenilirlik katsayılarının her iki desende de farklılaştığı tespit edilmiştir.

Çalışmanın sonuçlarına dayalı olarak aşağıdaki önerilere yer verilmiştir.

Veri sayısı aynı kalmak şartıyla puanlayıcıların birbirleriyle aynı ya da birbirlerinden farklı sayıda öğrenciyi puanlaması durumlarından elde edilen güvenilirlik katsayıları çok az farklılık göstermektedir. Dengelenmiş ve dengelenmemiş durumlar arasındaki bu farklılığın düşük olmasının nedeni, veri sayısının her iki desende de aynı tutulması gösterilebilir. Bir miktar veri kaybı olduğunda dengelenmemiş desende dengelenmiş desene göre daha yüksek güvenilirlik elde edilmektedir. Bu nedenle dengelenmemiş veri yapılarından veri indirgemesi yapılarak dengeli desenlerle genellenebilirlik kuramı analizleri yapılmak istendiğinde çok fazla veri kaybı olduğunda güvenirliliğin düştüğü unutulmamalıdır. Bu nedenle gerçek koşullarda dengelenmemiş veri yapısında olan durumlarda güvenilirlik hesaplanırken, dengelenmemiş desenlerden fazla veri kaybıyla dengeli yapılar oluşturarak güvenilirlik hesaplamak yerine, dengelenmemiş desenlerle analiz yapmak daha uygun olacaktır.

Puanlanan toplam öğrenci sayısı sabitken puanlayıcıların her birinin puanladığı öğrenci sayısı birbirlerine göre farklılaştıkça, bir başka deyişle puanlayıcıların puanladığı öğrenci sayıları arasındaki değişkenlik arttıkça G ve Phi katsayılarının düşme eğiliminde olduğu tespit edilmiştir. Bu nedenle OSCE sınavı İntramusküler Enjeksiyon yapma istasyonu ya da bu sınavın diğer istasyonlarında öğrencilerin performansları belirlenirken puanlayıcılar arasındaki öğrenci sayısı bakımından değişkenliğin fazla olmamasına dikkat edilmelidir. Ayrıca performansa dayalı güvenirliliğin araştırıldığı farklı eğitim araştırmalarında da veri doğası gereği dengelenmemiş durumda olduğunda değişkenin koşullarındaki gözlem sayılarına ait değişkenlik arttığında güvenirliliğin düşme eğiliminde olduğu unutulmamalıdır. Bu nedenle veri toplama aşamasında değişkenin koşullarına ait gözlemler olabildiğince birbirlerine yakın tutulmalıdır.

Yapılabilecek bir başka çalışmada da farklı performans durumlarında, farklı değişkenlik kaynakları kullanılarak dengelenmiş ve dengelenmemiş desenlerin karşılaştırılması araştırılabilir.

Kaynakça

- Arias, R. M. (2010). Performance Assessment. *Papeles del Psicologo*, 31(1), 85-96.
- Atılğan, H. (2008). Using Generalizability Theory to Assess The Score Reliability of The Special Ability Selection Examinations for Music Education Programmes in Higher Education. *International Journal of Research & Method in Education*, 31(1), 63-76.
- Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlog.
- Cronbach, L. J., Gleser, G. C., Nanda, H. & Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley.
- Deliceoğlu, G., Çıkrıkçı Demirtaşlı, N. (2012). Futbol Yetilerine İlişkin Dereceleme Ölçeğinin Güvenirliğinin Genellenebilirlik Kuramına ve Klasik Test Kuramına Dayalı Olarak Karşılaştırılması. Güvenirliklerinin Karşılaştırılması. *Spor Bilimleri Dergisi*, 23(1), 1-12.
- Elçin, M., Odabaşı, O., ve Sayek, İ. (2005). Yapılandırılmış Objektif Klinik Sınavlar. *Hacettepe Tıp Dergisi*, 36, 1-2.
- Güler, N., ve Gelbal, S. (2010). Studying Reliability of Open Ended Mathematics Items According to Classical Test Theory and Generalizability Theory. *Educational Sciences: Theory and Practice*, 10(2), 989-1019.
- Güler, N. (2011). Rastgele Veriler Üzerinde Genellenebilirlik Kuramı ve Klasik Test Kuramı'na Göre Güvenirliğin Karşılaştırılması. *Eğitim ve Bilim*, 36, 162: 225-234.
- Jeon, M. J., Lee, G., Hwang, J. W. & Kang, S. J. (2009). Estimating Reliability of School-Level Scores Using Multilevel and Generalizability theory Models. *Asia Pacific Education Rev.*, 10, 149-158.
- Kaufman, J. & Schering, A. (2007). *Analysis of Variance ANOVA*. Wiley Encyclopedia of Clinical Trials, John Wiley & Sons, Inc.
- Lee, G. & Frisbie, D. A. (1999). Estimating Reliability Under a Generalizability Theory Model for Test Scores Composed of Testlets. *Applied Measurement in Education*, 12 (3), 237 -255.
- Malhotra, M. K. & Sharma, S. (2008). Measurement Equivalence Using Generalizability Theory: An Examination of Manufacturing Flexibility Dimensions. *Decision Sciences*, 39(4), 643-669.
- Ødegård, A., Hagtvet, K. A. & Bjørkly, S. (2008). Applying Aspects of Generalizability Theory in Preliminary Validation of the Multifacet Interprofessional Collaboration Model (PINCOM). *International Journal of Integrated Care*, 8, 1-11.
- Sharma, S. & Weathers, D. (2003). Assessing Generalizability of Scales Used in Cross National Research. *International Journal of Research in Marketing*, 20, 287-295.
- Shavelson, J. R. & Webb, N. M. (1981). Generalizability Theory:1973-1980. *British Journal of Mathematical and Statistical Psychology*, 34, 133-166.
- Shavelson, J. R. & Webb, N. M. (1991). *Generalizability Theory: A Primer*. Newbury Park. CA: Sage Publications.
- Taşdelen, G. (2009). *Nedelsky ve Angoff Standart Belirleme Yöntemlerinin Genellenebilirlik Kuramı ile Karşılaştırılmasına İlişkin Bir Araştırma*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Webb, N. M., Shavelson, R. J. & Haertel, E. H. (2006). Reliability Coefficients and Generalizability Theory. *Handbook of Statistics*, 26, 81-124.
- Wei, X. & Haertel, E. (2011). The Effect of Ignoring Classroom-Level Variance in Estimating the Generalizability of School Mean Scores. *Educational Measurement: Issues and Practice*, 30(1), 13-22.
- Wiggins, G. (1993). Assessment: Authenticity, Context and Validity. *Phi Delta Kappan*, 75, 200-214.
- Wing, C. & Chiu, T. (2001). *Scoring Performance Assessments Based on Judgements: Generalizability Theory*. Kluwer Academic Publishers, Boston.

- Yelbođa, A. ve Tavřancıl, E. (2010). Klasik Test ve Genellenebilirlik Kuramı'na Gre Gvenirliđin Bir İř Performansı lçeđi zerinde İncelenmesi. *Kuram ve Uygulamada Eđitim Bilimleri*, 10(3), 1825-1854.
- Yılmaz Nalbantođlu, F. ve Gelbal, S. (2011). İletiřim Becerileri İstasyonu rneđinde Genellenebilirlik Kuramı'yla Farklı Desenlerin Karřılařtırılması. *Hacettepe niversitesi Eđitim Fakltesi Dergisi*, 41, 509-518.
- Yılmaz Nalbantođlu ve Bařusta Uzun (2012). "Genellenebilirlik Kuramıyla Dikiř Atma ve Alma Becerileri İstasyonu Gvenirliđinin Deđerlendirilmesi" III. Ulusal Eđitimde ve Psikolojide lçme ve Deđerlendirme Kongresi, Abant İzzet Baysal niversitesi, Bolu.