

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

GENELLENEBİLİRLİK KURAMINDA
DENGELENMİŞ VE DENGELENMEMİŞ DESENLERİN
KARŞILAŞTIRILMASI
-İntramuskuler Enjeksiyon Yapma İstasyonu Verileri
Üzerinde Bir Uygulama-

DOKTORA TEZİ

Funda Nalbantođlu Yılmaz

Ankara
Nisan, 2012

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI

GENELLENEBİLİRLİK KURAMINDA
DENGELENMİŞ VE DENGELENMEMİŞ DESENLERİN
KARŞILAŞTIRILMASI
-İntramuskuler Enjeksiyon Yapma İstasyonu Verileri
Üzerinde Bir Uygulama-

DOKTORA TEZİ

Funda Nalbantoğlu Yılmaz

Danışman: Prof. Dr. Ezel Tavşancıl

Ankara
Nisan, 2012

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne,

Bu alıřma j¼rimiz tarafından ¼lme ve Deđerlendirme Anabilim Dalında DOKTORA TEZİ olarak kabul edilmiřtir.

Bařkan
Prof. Dr. Nizamettin KO

¼ye
Prof. Dr. Ezel TAVŐANCIL (Danıřman)

¼ye
Prof. Dr. Selahattin GELBAL

¼ye
Prof. Dr. N¼khet IKRIKI DEMİRTAŐLI

¼ye
Do. Dr. Duygu ANIL

Onay
Yukarıdaki imzaların, adı geen ¼đretim ¼yelerine ait olduđunu onaylıyorum.

.../.../20..

Prof. Dr. Nejla KURUL
Enstit¼ M¼d¼r¼

TEŐEKKÜR

Yapıcı eleřtirileri ve önerileri ile alıřmama önemli katkılar sađlayan deđerli hocam ve danıřmanım Prof. Dr. Ezel TAVŐANCIL' a,

İtenliđi, önerileri ve yaklařımıyla desteđini hibir zaman esirgemeyen hocam Prof. Dr. Selahattin GELBAL'a, tezimi okuyarak deđerli önerileriyle destek veren hocalarım Prof. Dr. Nizamettin KO, Prof. Dr. Nkhet IKRIKI DEMİRTAŐLI ve Do. Dr. Duygu ANIL'a;

Arařtırma srecinde ve Yapılandırılmıř Objektif Klinik Sınav verilerini kullanmamda her trl kolaylıđı sađlayan Do. Dr. Melih ELİN'e, Arař. Gör. Bilge UZUN BAŐUSTA'ya;

Desteklerini her kořulda hissettiđim ve varlıklarıyla bana g veren eřim Kadri YILMAZ'a ve ANNEM'e;

Motivasyonum her dřtđnde beni harekete geiren ve yreklendiren, alıřma prensibiyle rnek aldıđım Do. Dr. Erdođan İEK'e;

Doktora eđitimim boyunca sađladıđı maddi katkıdan dolayı TBİTAK'a ve aldıđım eđitim boyunca đrettikleriyle yoluma ıřık tutan tm hocalarıma;

TEŐEKKÜR EDERİM.

ÖZET

GENELLENEBİLİRLİK KURAMINDA DENGELENMİŞ VE DENGELENMEMİŞ DESENLERİN KARŞILAŞTIRILMASI

- İntramuskuler Enjeksiyon Yapma İstasyonu Verileri Üzerinde Bir Uygulama-

Yılmaz Nalbantoğlu, Funda

Doktora, Ölçme ve Değerlendirme Anabilim Dalı

Tez Danışmanı: Prof. Dr. Ezel Tavşancıl

Nisan 2012, 96 sayfa

Araştırmanın amacı, Tıp Fakültesi Yapılandırılmış Objektif Klinik Sınav'daki 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 desenlerle yapılan genellenebilirlik kuramı sonuçlarını 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 değerlendirilmesinde ise sekiz puanlayıcı kullanılmıştır. Araştırmada, puanlayıcıların her birinin araştırmaya katılan öğrencilerden sadece bir kısmını 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.

Puanlayıcıların her birinin birbiriyle eşit (dengelenmiş) ya da birbirinden farklı (dengelenmemiş) sayıda öğrenciyi puanlamasıyla oluşturulmuş desenlerle yapılan G çalışmaları sonucunda kestirilen varyans değerleri her iki desende de benzerlik göstermektedir. Ayrıca senaryolara göre toplam veri sayısı aynı olmak şartıyla dengelenmiş ve dengelenmemiş (ö:p) x g desenleri ile kestirilen güvenilirlik katsayılarının her iki desende de birbirlerine yakın sonuçlar ürettiği tespit edilmiştir.



ABSTRACT

COMPARISON of BALANCED & UNBALANCED DESIGNS in GENERALIZABILITY THEORY

-An Application on the Data of Intramuscular Injection Station-

Yılmaz Nalbantođlu, Funda

Doctoral, Department of Measurement and Evaluation

Advisor: Prof. Dr. Ezel Tavşancıl

April 2012, 96 sayfa

This study aims to compare the results of the generalizability theory by employing the balanced and unbalanced designs that were formed by rating the groups of participant students, both equal and different in numbers, considering their performance on the same task provided that the number of the participating students stays the same.

The study group of the research are 240 first year students, who have taken the Intramuscular Injection Station in the Structured Objective Clinical Test of the Medicine Faculty of Hacettepe University. Eight raters was employed in assessing the performances of the students in the station. With each rater scoring the performance of certain students, the nested situation of the student (s) and rater (r) variables was examined while the same tasks (t), the (s:r) x t design matched with these variables were studied based on unbalanced and balanced situations.

The variance values retrieved from the G studies done by each rater's rating the students equal in number (balanced) and those unequal in number (unbalanced) were found to be parallel in both designs. Moreover, considering

the G and Phi coefficients obtained from the scenarios in the balanced and unbalanced (s:r) x t designs were observed to be close.



İÇİNDEKİLER

JÜRİ ÜYELERİNİN İMZA SAYFASI.....	i
TEŞEKKÜR.....	ii
ÖZET	iii
ABSTRACT.....	v
ÇİZELGELER LİSTESİ	ix
ŞEKİLLER LİSTESİ	xi
EKLER LİSTESİ.....	xii
BÖLÜM	1
1. GİRİŞ	1
Problem.....	1
Amaç.....	21
Önem	23
Sayıtlar	24
Sınırlılıklar	24
Tanımlar.....	25
Kısaltmalar	25
2. İLGİLİ ARAŞTIRMALAR	27
Yurt İçinde Yapılan Çalışmalar	27
Yurt Dışında Yapılan Çalışmalar.....	31
3. YÖNTEM	37
Araştırmanın Modeli.....	37
Çalışma Grubu.....	37
Veriler ve Toplanması	38
Verilerin Analizi	42

4. BULGULAR VE YORUMLAR	44
Dengelenmiş (ö:p) x g Deseninin Genellenebilirlik Kuramı Sonuçları.....	44
Dengelenmiş (ö:p) x g Deseniyle G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranları	44
Dengelenmiş (ö:p) x g Deseninde Puanlayıcı ve Öğrenci Sayılarının Artırılıp Azaltılması Senaryolarına Göre Yapılan Karar Çalışması	47
Dengelenmemiş (ö:p) x g Deseninin Genellenebilirlik Kuramı Sonuçları.....	50
Dengelenmemiş (ö:p) x g Deseniyle G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranları	50
Dengelenmemiş (ö:p) x g Deseninde Puanlayıcı ve Öğrenci Sayılarının Artırılıp Azaltılması Senaryolarına Göre Yapılan Karar Çalışması	53
Dengelenmiş ve Dengelenmemiş (ö:p) x g Desenlerinden Elde Edilen Genellenebilirlik Kuramı Sonuçlarının Karşılaştırılması.....	57
Her İki Desenden Elde Edilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranlarının Karşılaştırılması.....	57
Her İki Desende Puanlayıcı ve Öğrenci Sayılarının Artırılıp Azaltılmasıyla Yapılan Karar Çalışmalarının Karşılaştırılması	60
5. SONUÇ VE ÖNERİLER.....	65
Sonuçlar	65
Öneriler	68
KAYNAKÇA	71
EKLER	82

ÇİZELGELER LİSTESİ

Çizelge 1. Dengelenmiş Desen İçin Varyans Bileşenlerinin Kestirilmesi	17
Çizelge 2. Dengelenmemiş Desen İçin Varyans Bileşenlerinin Kestirilmesi...	19
Çizelge 3. (ö.p) x g Deseni Dengelenmiş Veri Yapısı Örneği	39
Çizelge 4. (ö.p) x g Deseni Dengelenmemiş Veri Yapısı Örneği	40
Çizelge 5. Araştırma Kapsamında Kullanılan Dengelenmiş ve Dengelenmemiş (ö:p) x g Desenlerine Ait Betimsel İstatistikler	41
Çizelge 6. Dengelenmiş (ö:p) x g Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranları.....	45
Çizelge 7. Dengelenmiş (ö:p) x g Desenine Ait K Çalışması İle Puanlayıcı ve Öğrenci Sayılarının Değiştirilmesi Senaryolarına Göre Kestirilen G ve Phi Katsayıları	47
Çizelge 8. Dengelenmemiş (ö:p) x g Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranları.....	51
Çizelge 9. Dengelenmemiş (ö:p) x g Desenine Ait K Çalışması İle Puanlayıcı ve Öğrenci Sayılarının Değiştirilmesi Senaryolarına Göre Kestirilen G ve Phi Katsayıları	54

Çizelge 10. Dengelenmiş ve Dengelenmemiş (ö:p) x g Desenlerinden Elde Edilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranları..... 58

Çizelge 11. Dengelenmiş ve Dengelenmemiş (ö:p) x g Desenleri İle Senaryolara Göre Yapılan K Çalışmalarından Elde Edilen G ve Phi Katsayılarının Karşılaştırılması 61



ŞEKİLLER LİSTESİ

Şekil 1. (ö:p) x g Deseni Varyans Kaynakları Venn Şeması 8



EKLER

Ek 1. Tıp Eğitimi ve Bilişimi Anabilim Dalı İntramuskuler Enjeksiyon Yapma Becerisi Değerlendirme Formu	82
Ek 2. Dengelenmiş (ö:p) x g Deseni Syntax ve Analiz Sonuçları	83
Ek 3. Dengelenmemiş (ö:p) x g Deseni Syntax ve Analiz Sonuçları	90

BÖLÜM I

GİRİŞ

Bu bölümde, araştırma problemi, araştırmanın amacı, önemi, sayıtları, sınırlılıkları ile araştırmayla ilgili tanımlar ve kısaltmalar açıklanmıştır.

Problem

Bireyler hakkında alınan kararların önemli olduğu günümüzde, ölçme ve değerlendirme büyük önem taşımaktadır. Çeşitli amaçlarla bireyler hakkında alınan kararların doğruluğu ve isabetliliği için ölçme ve değerlendirme işlemlerinin güvenilirliğinin ve geçerliğinin olabildiğince yüksek olması beklenmektedir. Bu nedenle, ölçme ve değerlendirme sürecinde amaç ile araç arasındaki bağlantı dikkati çekmekte; ölçülmek istenen özelliklerin doğru ve uygun ölçme ve değerlendirme yöntemleriyle ölçülmesi gerekmektedir. Uygun ölçme ve değerlendirme yönteminin seçilmesinde ise ölçülecek davranışın bilişsel, duyuşsal ve devinişsel alanlardan hangisinin niteliklerinde olduğuna karar vermek, hedef durum ile ilgili soru tipini belirlemede önemli bir kriterdir.

Eğitimin her kademesinde ve personel seçiminde, zihin ve kas koordinasyonu gerektiren becerilerin değerlendirilmesinde performans değerlendirme oldukça önem taşımaktadır. Özellikle bir meslek grubuna ait becerilerin kazandırıldığı yükseköğretim programlarında performans değerlendirme, kullanılması gereken önemli değerlendirme yöntemlerindedir. Yükseköğretimde yapılan ölçme ve değerlendirme çalışmalarının öğrencilerin üst düzey düşünme becerilerini yoklamaya yönelik olması önemlidir. Bu nedenle bir meslek grubuna ait bilgi ve becerilerin kazandırıldığı öğrencilerin

problem çözüme, eleştirel düşünme, akıl yürütme ve yaratıcı düşünme gibi üst düzey zihinsel becerilerinin değerlendirilmesi gerekmektedir. Öğrencilerin üst düzey zihinsel becerilerinin değerlendirilmesinde ise performans değerlendirme dikkati çekmektedir.

Performans, üst düzey zihinsel süreç gerektiren bir görevi ya da etkinliği yerine getirirken ortaya konan çaba ve bu çaba sonunda ortaya çıkan ürün olarak adlandırılır (Kutlu, Doğan ve Karakaya, 2010). Performans değerlendirme ise, belirlenmiş bir görevi yürütmek için gerekli olan bilgi ve becerilerin yapay ya da gerçek yaşam durumlarında uygulanmasını içeren bir değerlendirmedir (Arias, 2010; Palm, 2008). Arter ve Stiggins (1992)'e göre ise performans değerlendirme, gözlem ve kanıya dayalı değerlendirme olarak tanımlanmaktadır.

Performans, somut bir ürünle ve/veya gözlenebilen bir süreç (beceri) ile sonuçlanmaktadır. Böylelikle performans değerlendirmede öğrencilerin karmaşık bir görevi yerine getirmesi ile elde edilen ürün ve/veya süreç değerlendirilir.

Performans değerlendirmede, öğrenci kendisine verilen problemi çözebilmek ya da ilgili görevi yapabilmek için ön bilgisini kullanmaya, araştırma yapmaya ve sorgulamaya ihtiyaç duyar (Bekiroğlu, 2008). Böylece performans değerlendirme, öğrencilerin bir durum karşısında edindiği temel bilgi ve becerilerini kullanmayı gerektirir. Performans değerlendirmenin klasik değerlendirme yöntemlerinden farkı, öğrencinin edindiği bilgi ve becerileri kullanıp bu bilgi ve becerileriyle yeni ürünler üretmesidir (Airasian, 1994). Performans değerlendirmeyle öğrenci edindiği bilgileri yeniden yorumlayarak özgünlüğünü ve yaratıcılığını ortaya çıkartır. Böylelikle performans değerlendirme, öğrencilerin üst düzey düşünme becerilerine de hizmet ederek öğrencilerin eleştirel düşüncelerini ve yaratıcılıklarının gelişmesini sağlamaktadır (Moss, 1992). Performans değerlendirmenin sağladığı bir başka avantaj ise öğrencilere neler öğrendiklerini gösterebilme fırsatı sunmasıdır. Bu sayede öğrencinin güçlü ve zayıf yanlarının hem kendisi hem de öğretmeni tarafından fark edilmesi sağlanır (Brualdi, 1998). Öğrencilerin

bilgiyi nasıl yapılandırdıkları belirlenerek onlara daha iyi bir değerlendirme olanağı sunulur (Khattri, Reeve ve Kane, 1998).

Performans değerlendirme özünde bir görevi ya da süreci yürütme ve sonuçlandırmayı gerekli kılar (Wiggins, 1993). Bu tür değerlendirmeler 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).

Öğrencilerin ilgili becerileri nasıl yaptıklarının değerlendirildiği performans sınavlarından biri de “Yapılandırılmış Objektif Klinik Sınav” (Objective Structure Clinical Examination) kısaca YOKS/OSCE’dir. Bu sınav, birden fazla istasyondan oluşan ve her bir istasyonda farklı klinik becerilerin puanlayıcılar tarafından değerlendirildiği bir performans sınavıdır. Elde edilen sınav sonuçları tıp fakültesi öğrencilerinin klinik becerilerine dönük önemli bilgiler elde edilmesini sağlar. Bu sınav, 1979 yılında yaygın kullanımı ile kabul edilmiş olup başta ABD ve Kanada olmak üzere Türkiye’de de tıp fakültesi öğrencilerine uygulanan ve standart becerileri kazandırmaya dönük klinik bir sınavdır (Clauser, Harik, Margolis, Mee ve Swygert, 2008; Harden ve Gleeson, 1979).

OSCE, Hacettepe Üniversitesi Tıp Fakültesi Tıp Eğitimi ve Bilişimi Anabilim Dalı tarafından 2004-2005 öğretim yılından itibaren iyi hekimlik uygulamaları programının yılsonu değerlendirmelerinde kullanılmaya başlanmıştır (Elçin, Odabaşı ve Sayek, 2005). İyi hekimlik uygulamalarıyla öğrencilere iyi bir hekim olabilmenin gereksinimleri doğrultusunda iletişim becerileri, mesleki beceriler ve muayene becerileri kazandırılarak, etik değerler, klinik ziyaretler ve kanıta dayalı tıp çalışmalarıyla desteklenen bir program sunulmaktadır (Akyüz ve diğerleri, 2011).

İyi hekimlik uygulamaları programında yer alan mesleki beceriler alt bölümünde temel beceriler, muayene becerileri, klinik ve laboratuvar becerileri ele alınmaktadır (Akyüz ve diğerleri, 2011). İyi hekimlik uygulamalarıyla doktor

adaylarına kazandırılmak istenen bu mesleki beceriler, uygulamalı olarak OSCE ile değerlendirilmektedir.

Araştırma kapsamında, iyi hekimlik uygulamaları programının mesleki beceriler alt bölümünden “İntramuskuler Enjeksiyon” yapma becerilerinin değerlendirildiği istasyon ele alınmıştır. “İntramuskuler Enjeksiyon” yapma istasyonu öğrencilerin gerçek yaşam koşullarında karşılaşılabilecekleri enjeksiyon yapma mesleki becerilerine ait performans durumlarının değerlendirilmesini sağlar.

OSCE sınavı da tüm performans değerlendirme çalışmalarında olduğu gibi çeşitli faktörlerden etkilenmektedir. Bu faktörler sınavın yapıldığı ortamdaki, 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 doğruluğu ve isabetliliği değişmektedir. Bu nedenle, elde edilen performans puanlarını kullanmadan önce puanların güvenilirliklerinin araştırılması önem taşımaktadır. Performansın ölçülmesinde güvenilirlik daha çok klasik test kuramı ve genellenebilirlik kuramına dayalı yöntemlerle çalışılmaktadır (Briesch, Chafaouleas ve Riley-Tillman, 2010; Kreiter ve Ferguson, 2001; Lane ve Sabers, 1989; Lei, Smith ve Suen, 2007; Touchie, Humphrey, Ainslie, Myers ve Wood, 2010).

Tarihsel olarak güvenilirlik, eğitim ve psikolojide klasik test kuramı ile ele alınmıştır (Feldt & Brennan, 1989 akt Brennan, 2000). Klasik test kuramına göre güvenilirlik, gözlenen puanların gerçek puanları ne derece doğru yansıttığıdır (Crocker ve Algina, 1986). Klasik test kuramında gözlenen puan (X), gerçek puan (T) ve hata puanından (E) oluşmakta ve Eşitlik 1'deki gibi gösterilmektedir.

$$X = T + E \quad (1)$$

Gerçek puanın hata puanı ile ilişkisiz olduğu varsayılarak, gözlenen puan varyansı (σ^2_X), gerçek puan varyansı (σ^2_T) ve hata puanı varyansının (σ^2_E) toplamı olarak Eşitlik 2'deki gibi gösterilmektedir.

$$\sigma^2_X = \sigma^2_T + \sigma^2_E \quad (2)$$

Klasik test kuramında güvenilirlik katsayısı, gerçek puan varyansının gözlenen puan varyansına oranıdır. Böylece klasik test kuramında güvenilirlik katsayısı Eşitlik 3'teki gibi gösterilmektedir.

$$P^2_{XT} = \sigma^2_T / \sigma^2_X = \sigma^2_T / (\sigma^2_T + \sigma^2_E) \quad (3)$$

Eşitlik 3'te verilen bu güvenilirlik katsayısı, bireylerin gerçek puanları bilinmediğinden pratik değildir (Webb, Shavelson ve Haertel, 2006). Bu nedenle güvenilirlik katsayısı çeşitli varsayımlara dayalı farklı yöntemler aracılığıyla dolaylı olarak elde edilmektedir. Dolayısıyla, farklı kaynaklardan gelen hataların ölçme sonucuna karışabildiği durumlarda paralel testler güvenilirliği, test tekrar test, eşdeğer formlar, iç tutarlık ve puanlayıcılar arası güvenilirlik gibi çeşitli güvenilirlik yöntemleriyle güvenilirliği değerlendirmek gerekir. Diğer bir deyişle, klasik test kuramında her hata kaynağına göre en az bir güvenilirlik tahmini vardır ve bu güvenilirlikler hata kaynağına bağlı olarak değişir (Eason, 1989; Suen ve Lei, 2007). Görüldüğü gibi bu yaklaşım pratikte oldukça zahmetlidir.

Genellenebilirlik kuramına göre güvenilirlik, genellenebilirlik kavramı ile yer değiştirmiştir. Genellenebilirlik kuramında ölçme sonuçlarının gerçek puana ne ölçüde uygun olduğundan çok ölçme sonuçlarının genelleme yapılacak evrene ne ölçüde uygun olduğu üzerinde durulur (Şencan, 2005).

Klasik test kuramındaki gerçek puan varyansının yerini genellenebilirlik kuramında evren puanı varyansı almıştır. Evren, araştırmacının genelleme yapmak istediği ana küttedir (Brennan 2001). Böylelikle güvenilirlik anlamındaki genellenebilirlik katsayısı, evren puanı (universe score) varyansının gözlenen puan varyansına oranıdır. Gözlenen puan varyansı da

evren puanının varyansı ve bağıl hata varyansının toplamına eşittir (Brennan, 2001; Shavelson ve Webb, 1991).

Klasik test kuramında test puanlarının varyansı, bireyin gerçek yeteneğinin varyansı ve hata varyansından oluşmaktadır. Klasik test kuramında tek bir gerçek puan ve ayırt edilmemiş bir hata terimi olmasına rağmen ölçme sonuçlarına karışan hatanın birden fazla kaynaktan gelebileceği ileri sürülmüştür (Cronbach, Gleser, Nanda ve Rajaratnam, 1972; Cronbach, Rajaratnam ve Gleser, 1963 akt Kraiger, 1989). Oysa 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 ve diğerleri, 1972; Gillmore, Kane ve Naccarato, 1978; Nocera, Ferlazzo ve Borghi, 2001; Shavelson ve Webb 1991). Böylece klasik test kuramının tek bir hata kaynağını ele alması düşünüldüğünde genellenebilirlik kuramıyla birçok hata ve bu hataların etkileşimleri dikkate alınarak daha kapsamlı ve gerçekçi hata betimlemesi yapılabilir.

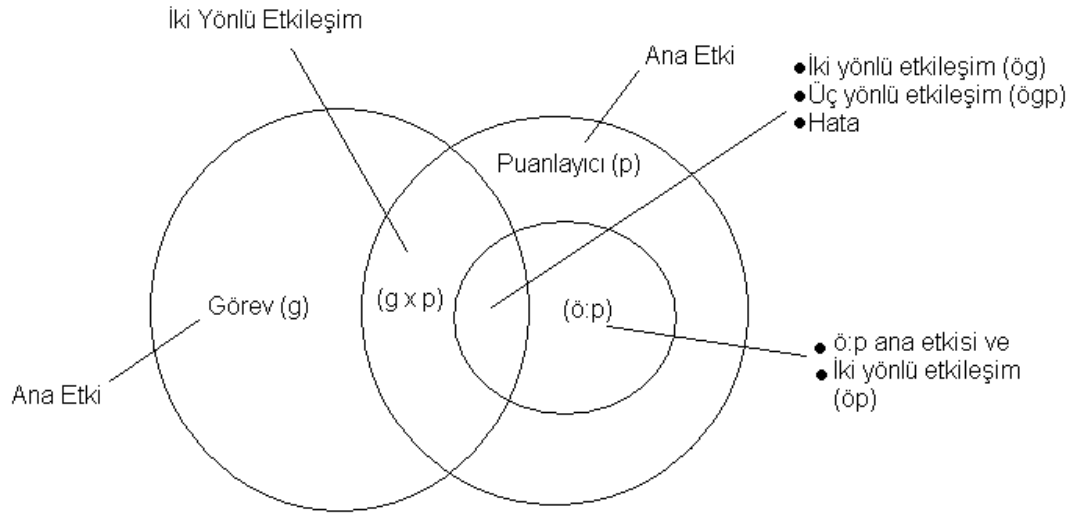
Genellenebilirlik kuramı, gözlenen puanlardaki tutarsızlık kaynaklarının miktarının belirlenmesini ve davranışın ölçülmesinde güvenirliliğin değerlendirilmesini, araştırılmasını sağlayan istatistiksel bir kuramdır (Brennan, 2001; Cronbach ve diğerleri, 1972; Shavelson ve Webb, 1991).

Genellenebilirlik kuramı Cronbach, Gleser, Nanda ve Rajaratnam (1972) tarafından klasik test kuramına alternatif olarak geliştirilmiştir (Brennan, 2001; Shavelson ve Webb, 1991). Geniş uygulama alanına sahip olan genellenebilirlik kuramının esas yapısı varyans analizine dayanmaktadır. Test durumuna ait her bir değişkenin varyans bileşenleri varyans analiziyle kestirilir. Kavramsal olarak genellenebilirlik kuramı, klasik test kuramının çok yönlü bir uzantısı olarak kabul edilebilir ve klasik test kuramı ile varyans analizinin birleşimi olarak görülebilir (Brennan, 2000; Suen ve Lei, 2007).

Gözlenen puan varyansı ise;

$$\hat{\sigma}^2(X_{(öpg)}) = \hat{\sigma}^2_g + \hat{\sigma}^2_p + \hat{\sigma}^2_{ö:p} + \hat{\sigma}^2_{gp} + \hat{\sigma}^2_{gö:p,e} \quad (5)$$

Eşitlik 5'teki gibi gösterilmektedir. Eşitlik 5'te verilen gözlenen puan varyansı toplam varyans olarak da adlandırılır ve G çalışmasında elde edilen varyans bileşenlerinin toplamıdır (Brennan, 2000). Toplam varyansa ait varyans bileşenleri Şekil 1'deki gibi şematize edilebilir.



Şekil 1. (ö:p) x g Deseni Varyans Kaynakları Venn Şeması

Eşitlik 5 ve Şekil 1'de de görülebileceği gibi (ö:p) x g desenine ait toplam varyans ana ve ortak etki olmak üzere beş varyans bileşenine ayrılmaktadır. Bu varyans bileşenlerinde "ö" öğrencileri, "g" görevleri, "p" de puanlayıcıları göstermektedir. Her bir varyans kaynağı için varyans analiziyle hesaplanan değerler araştırmacı için bilgi sağlamaktadır. Görevlere ait varyans bileşeni ($\sigma^2(g)$), bazı görevlerin diğerlerine göre zorluk-kolaylık bakımından farklılaşıp farklılaşmadığını, puanlayıcılara ait varyans bileşeni ($\sigma^2(p)$), puanlayıcılar arasında farklılık olup olmadığını, puanlayıcı değişkenliğinin öğrencilerin performansını etkileyip etkilemediğini

göstermektedir. Öğrencilerin puanlayıcılarla yuvalandığı $\sigma^2(\ddot{o}:p)$ varyans bileşeni, ölçülen beceri bakımından öğrencilerin birbirlerinden farklılık gösterip göstermediğini, puanlayıcıların davranışlarının bir öğrenciden diğerine değişip değişmediğini göstermektedir. Görev x puanlayıcı etkisine ait varyans bileşeni ($\sigma^2(gp)$), görev-puanlayıcı etkisinden kaynaklı farklılıkların olup olmadığını, puanlayıcıların bir görevden diğerine verdiği puanlardaki kararlılıklarını gösterir (Shavelson ve Webb, 1991). Artık etkisine ait varyans bileşeni ($\sigma^2(\ddot{o}gp)$) ise öğrenci-görev ortak etkileşimi, üç yönlü etkileşim (öğrenci-görev-puanlayıcı) ve/veya diğer değişkenlik kaynaklarına ait bilgi içerir (Brennan, 2001; Shavelson ve Webb, 1991).

Karar (D) Çalışması

Karar çalışmaları, hata varyansının gerçek puan üzerindeki etkisinin azaltılabilmesi için farklı senaryo durumlarına göre araştırmacıların incelemeler yapmasına olanak sağlar. D çalışması, kabul edilebilir hata varyansı ile farklı senaryolara göre güvenilirlik katsayısı kestirimleri yapmada kullanılır (Brennan, 2000). Bu açıdan D çalışmasının amacı, G çalışması sonuçlarına dayanılarak farklı senaryoların etkililiğini karşılaştırarak hatayı en aza indirmektir (Pedersen, Hagtvet ve Karterud, 2007). Böylece hata varyansının azaltılıp güvenirliliğin artırılabilmesi için araştırmacıya alternatif yollar sunulur (Shavelson ve Webb, 1991). Araştırmacı gelecek çalışmalar hakkında ön bilgi sahibi olur. Bu açıdan genellenebilirlik kuramıyla farklı senaryo durumlarına göre karar çalışması ile güvenilirlik kestirimlerinin yapılabilmesi klasik test kuramına göre avantajlı bir durumdur.

Araştırmacı bir D çalışması planlarken genellemek istediği değişken ve koşulları için genelleme evreni (universe of generalization) tanımlar (Webb, Shavelson ve Haertel, 2006). Genelleme evreni, araştırmacının çalışmada dikkate aldığı, genellemek istediği koşullar setidir. Araştırmada belirlenen ölçme konusu (object of measurement) hariç diğer değişkenler genelleme evreni olur (Cardinet, Tourneur ve Allal, 1976). Genelleme evreni, araştırmacının amacına bağlı olarak ölçme konusunu (object of

measurement) ve evren puanı varyansını belirleyerek genellenebilirlik katsayısını hesaplamasına yardımcı olur (Brennan, 2001; Cardinet, Tourneur ve Allal, 1981; Durvasula, Netemeyer, Andrews ve Lysonski, 2006). Genel olarak ölçmenin konusu bireyler olmakla birlikte sınıf, okul, ülke, grup, öğretmen ve puanlayıcıların ölçme konusu olarak kullanıldığı çalışmalar da bulunmaktadır (Barnes ve Barnes, 1993; Brennan, Yin ve Kane, 2003; Cardinet ve diğerleri, 1976; Chang ve Hocevar, 2000; Cronbach, Linn, Brennan ve Haertel, 1997; Dolmans, Wolfhagen, Schmidt ve Vleuten, 1994; Fitzpatrick, Lee ve Gao, 2001; Foy, 2005; Gillmore, Kane ve Naccarato, 1978).

Genellenebilirlik çalışmasında elde edilen varyans kestirimleri D çalışmasında kullanılır. Bu varyans kestirimleri yardımıyla da bağıl ve mutlak hata varyansları hesaplanarak genellenebilirlik (E_p^2) ve Phi (Φ) katsayıları olmak üzere iki tip güvenilirlik katsayısı hesaplanır (Brennan, 2000).

Marcoulides (2000), klasik test kuramına göre genellenebilirlik kuramını kullanmanın bir avantajını genellenebilirlik kuramıyla iki tür hatanın ayırt edilebilmesi olarak göstermiştir. Bu bağlamda genellenebilirlik kuramında hata varyansı bağıl (relative) ve mutlak (absolute) hata varyansı olmak üzere ikiye ayrılır (Brannan, 2001; Shavelson ve Webb, 1991).

Bağıl hata, bireyin gözlenen sapma puanı ile evren sapma puanı arasındaki farktır (Brennan, 2000, 2001). Bağıl hata varyansı bağıl değerlendirmede kullanılır. Eğer araştırmacı öğrencilerin performanslarını birbirleriyle karşılaştırarak değerlendirmek istiyorsa elde edilecek güvenilirlik katsayısında bağıl hata varyansını kullanır. Bağıl hata varyansı " $\sigma^2(\delta)$ " şeklinde gösterilir (Brennan, 2001).

Bağıl hata varyansı evren puanının ortak etkilerini içerir (Shavelson ve Webb, 1991). Böylece Eşitlik 5 ve Şekil 1'de varyans bileşenlerinin verildiği ve öğrencilerin farklı puanlayıcılar tarafından aynı görevler doğrultusunda puanlanmasıyla görevlerin ölçme konusu (object of measurement) ve tüm değişkenlerin tesadüfî olarak belirlendiği (ö:p) x g deseni (ö: öğrenci ve p:

puanlayıcı ve g: görev olmak üzere) için hesaplanan bağıl hata varyansı Eşitlik 6'daki gibi gösterilmektedir.

$$\begin{aligned}\sigma^2(\delta) &= \sigma^2(gP) + \sigma^2(g\ddot{O}:P) \\ &= \frac{\sigma_{gp}^2}{n_p} + \frac{\sigma_{g\ddot{O}:p}^2}{n_{\ddot{O}}n_p}\end{aligned}\quad (6)$$

Mutlak hata, bireyin gözlenen puanı ile evren puanı arasındaki farktır ve " $\sigma^2(\Delta)$ " şeklinde gösterilir (Brennan, 2000, 2001). Mutlak hata varyansı mutlak değerlendirmelerde kullanılır. Böylece öğrencilerdeki değişim gruba bağlı olmadan mutlak bir ölçüte göre belirlenir.

Mutlak hata varyansı, evren puanı varyansı hariç desendeki diğer tüm varyansların toplamıdır (Brennan, 2000; Webb ve Shavelson, 1981). Böylece yukarıda verilen aynı örneğe ait mutlak hata varyansı ise Eşitlik 7'deki gibi gösterilmektedir.

$$\begin{aligned}\sigma^2(\Delta) &= \sigma^2(P) + \sigma^2(gP) + \sigma^2(\ddot{O}:P) + \sigma^2(g\ddot{O}:P) \\ &= \frac{\sigma_p^2}{n_p} + \frac{\sigma_{gp}^2}{n_p} + \frac{\sigma_{\ddot{O}:p}^2}{n_{\ddot{O}}n_p} + \frac{\sigma_{g\ddot{O}:p}^2}{n_{\ddot{O}}n_p}\end{aligned}\quad (7)$$

Genellenebilirlik kuramında bağıl ve mutlak değerlendirmeler için bir ayırım söz konusudur. Böylece genellenebilirlik kuramı ile bağıl ve mutlak değerlendirmeler için bağıl ve mutlak hata varyanslarına göre genellenebilirlik (G) ve Phi katsayısı olmak üzere iki tür katsayı hesaplanabilir (Brennan, 2001). Bu iki katsayının klasik test kuramındaki güvenilirlik katsayısı gibi alınan kararlarda önemi vardır (Wing ve Chiu, 1999). G ve Phi katsayıları ise aşağıda ayrıntılı olarak açıklanmıştır.

Genellenebilirlik Katsayısı (Ep^2)

Genellenebilirlik katsayısı, klasik test kuramındaki gerçek puan varyansının gözlenen puan varyansına oranı olan güvenilirlik katsayısıyla benzerlik göstermektedir (Brennan, 2001; Rentz, 1987; Shavelson ve Webb, 1981). Ayrıca klasik test kuramındaki gerçek puan, genellenebilirlik kuramındaki evren puanına (universe score) benzemektedir (Brennan, 2000, 2001). Genellenebilirlik katsayısı, evren puanı varyansının gözlenen puan varyansına oranıdır (Shavelson ve Webb, 1981). Gözlenen puan varyansı da evren puanı varyansı ve bağıl hata varyansının ($\sigma^2(\delta)$) toplamına eşittir. Böylece genellenebilirlik katsayısı Eşitlik 8'de gösterildiği gibi hesaplanmaktadır (Brennan, 2001; Shavelson ve Webb, 1991).

$$Ep^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)} \quad (8)$$

Öğrencilerin farklı puanlayıcılar tarafından aynı görevler doğrultusunda dönüşümlü olarak puanlanmasıyla görevlerin ölçmenin konusu ve tüm değişkenlerin tesadüfi olarak belirlendiği (ö:p) x g deseninde evren puanı görevler, puanlayıcı ve öğrenciler genelleme evreni olarak alınabilir. Böylece (ö:p) x g deseninde evren puanı ve Eşitlik 6'da elde edilen bağıl hata varyansı Eşitlik 8'de yerine konulduğunda kestirilen genellenebilirlik katsayısı aşağıdaki gibi olmaktadır.

$$Ep^2 = \frac{\sigma^2(g)}{\sigma^2(g) + \frac{\sigma^2(gp)}{n_p} + \frac{\sigma^2(g\ddot{o}:p)}{n_{\ddot{o}}n_p}} \quad (9)$$

Eşitlik 9'daki gibi kestirilen genellenebilirlik katsayısı 0 ile 1 arasında değerler almaktadır.

Phi Katsayısı (Φ)

Mutlak değerlendirmelerde kullanılan güvenilirlik katsayısı ise phi katsayısıdır (Brennan ve Kane, 1977). Phi katsayısı, evren puanı varyansının gözlenen puan varyansına oranıdır. Gözlenen puan varyansı da evren puanı varyansı ve mutlak hata varyansının ($\sigma^2(\Delta)$) toplamına eşittir. Böylece phi katsayısı Eşitlik 10'daki gibi gösterilir (Brennan, 2001; Shavelson ve Webb, 1991).

$$\Phi = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\Delta)} \quad (10)$$

Öğrencilerin puanlayıcılarla yuvalandığı (ö:p) x g deseni örneğinde evren puanı ve mutlak hata varyansı yerine konulduğunda elde edilen phi katsayısı ise Eşitlik 11'deki gibi olmaktadır.

$$\Phi = \frac{\sigma^2(g)}{\sigma^2(g) + \frac{\sigma^2(p)}{n_p} + \frac{\sigma^2(gp)}{n_p} + \frac{\sigma^2(\ddot{o}:p)}{n_{\ddot{o}}n_p} + \frac{\sigma^2(g\ddot{o}:p)}{n_{\ddot{o}}n_p}} \quad (11)$$

Eşitlik 8 ve 10'da görüldüğü gibi genellenebilirlik katsayısı ve phi katsayısı arasındaki temel fark genellenebilirlik katsayısında bağıl hata varyansı kullanılırken, phi katsayısında mutlak hata varyansının kullanılmasıdır. Ayrıca, desenlere göre elde edilen mutlak hata varyansının bağıl hata varyansından büyük ($\sigma^2(\delta) \leq \sigma^2(\Delta)$) olduğu görülmektedir. Böylece aynı desene göre hesaplanacak olan G katsayısı her zaman phi katsayısından daha yüksek olmaktadır (Brennan, 2001).

Araştırmada kullanılan değişkenlere ait koşullar, araştırmacının ilgisine ve amacına bağlı olarak sabit (fixed) ve tesadüfî (random) olabilir (Brennan, 2000, 2001). Sabit değişkenin tüm koşulları evrende sabittir. Böylece

araştırmacı koşulları evrene genellemek istemez (Brennan, 2001). Tesadüfi değişkenin koşulları ise olası gözlemler evreninden seçilir ve araştırmacı örneklemin ötesinde genelleneme yapabilir. Örneğin; araştırmada kullanılan puanlayıcılar tesadüfi seçildiğinde sonuçlar benzer diğer puanlayıcılar için genellenebilir (Suen ve Lei, 2007). Eğer araştırmacı araştırmasında kullandığı puanlayıcıları sabit olarak seçerse, elde ettiği sonuçları diğer puanlayıcılar için genellemek istemez. Örneğin; her yıl uygulaması olan bir sınavda hep aynı puanlayıcılar kullanılmasına rağmen farklı grup soruların kullanılabilmesi bir durumda puanlayıcılar sabit, sorular ise tesadüfi olabilir.

Değişkenlerin koşullarının sabit veya tesadüfi seçilmesi güvenilirlik kestirimlerini etkilemektedir. Bu durum yukarıda görevlerin ölçme konusu ve değişkenlerin tesadüfi olarak belirlendiği (ö:p) x g deseninde, puanlayıcılar sabit olarak ele alındığında kestirilecek genellenebilirlik ve Phi katsayısı üzerinde aşağıda gösterilmiştir. Puanlayıcıların sabit olması durumunda genellenebilirlik katsayısı Eşitlik 12'deki gibi olmaktadır.

$$E\rho^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)} \quad (12)$$

$$\sigma^2(\tau) = \sigma^2(g) + \frac{\sigma^2(gp)}{n_p}$$

$$\sigma^2(\delta) = \frac{\sigma^2(g\ddot{o}:p)}{n_{\ddot{o}}n_p}$$

Aynı durumda Phi katsayısı ise Eşitlik 13'deki gibi olmaktadır.

$$\Phi = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\Delta)} \quad (13)$$

$$\sigma^2(\tau) = \sigma^2(g) + \frac{\sigma^2(gp)}{n_p}$$

$$\sigma^2(\Delta) = \frac{\sigma^2(\ddot{o}:p)}{n_{\ddot{o}}n_p} + \frac{\sigma^2(g\ddot{o}:p)}{n_{\ddot{o}}n_p}$$

Puanlayıcıların sabit olarak ele alınmasıyla hesaplanan Eşitlik 12 ve Eşitlik 13'teki genellenebilirlik ve Phi katsayılarında da görüldüğü gibi puanlayıcılar tesadüfi olarak ele alındığında hesaplanan Eşitlik 9 ve Eşitlik 11'deki genellenebilirlik ve Phi katsayıları arasında farklılıklar bulunmaktadır. Puanlayıcılar sabit olarak alındığında evren puanı varyansı artmakta, hata varyansları azalmaktadır. Böylece puanlayıcı değişkeni sabit olarak ele alındığında G ve Phi katsayıları puanlayıcıların tesadüfi olarak belirlenmesine göre yükselmektedir.

Sabit veya tesadüfi değişkenler kullanarak araştırma amacı doğrultusunda çalışma desenleri oluşturulur. Bu çalışma desenleri ile ilgili temel bir ayrım ise, değişkenlerin çapraz (x) ya da yuvalanmış (:) olarak tasarlanabilmesindedir. Değişkenlerin tamamının çaprazlandığı desende, bir değişkenin bütün koşulları diğer değişkenin bütün koşullarıyla örtüşmekte ve değişkenin bütün koşullarına ait puan elde edilmektedir (Mushquash ve O'connor, 2006). Örneğin; puanlayıcıların hepsinin bütün bireyleri aynı görevler doğrultusunda puanlaması; puanlayıcı, görev ve bireylerin çaprazlanmasıdır. Yuvalanmış desende ise bir değişkenin olası koşulları, diğer değişkenin olası koşullarının hepsi ile örtüşmemektedir. Örneğin; görevler puanlayıcılar ile yuvalanmış ise her bir puanlayıcı farklı görevleri puanlamaktadır. Tüm değişkenlerin çapraz ya da yuvalanmasıyla desenler oluşturulabilmekle birlikte değişkenlerin birbirleriyle yuvalanmış ve çapraz olabildiği karışık (mixed) desenler de oluşturulabilmektedir. Bununla birlikte tümüyle çapraz tasarımlar yuvalanmış desenlere göre daha çok gözlem sayısı gerektirmektedir.

Değişkenlerin birbirleriyle çapraz ya da yuvalanmış şekilde tasarlanarak oluşturulduğu desenler, araştırma verilerine göre kimi zaman dengelenmiş (balanced) kimi zaman da dengelenmemiş (unbalanced) olmaktadır. Tüm değişkenlerin çapraz tasarlandığı desenlerde, değişkenlerin koşullarına ait veri kaybı (missing value) olmadıkça dengelenmiş desenler oluşturulmaktadır. Fakat değişkenlerin birbirleriyle yuvalandığı desenlerde hem veri kaybı, hem de gerçek durum koşulları nedeniyle dengelenmemiş desenler ortaya çıkmaktadır (Brennan, 1992). Dengelenmiş ve dengelenmemiş desenlere ilişkin açıklamalar ise aşağıda sunulmuştur.

Dengelenmiş ve Dengelenmemiş Desenler

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 tasarım yine dengeli bir tasarımdır.

Eşitlik 4 ve 5'te öğrencilerin bir performans durumuna ait becerilerinin değerlendirilmesinde belli sayıdaki öğrencilerin farklı puanlayıcılar tarafından dönüşümlü olarak aynı görevler doğrultusunda puanlanmasıyla oluşturulmuş (ö:p) x g (ö: öğrenci, p: puanlayıcı ve g: görev olmak üzere) desenine ait gözlenen puan ve gözlenen puanların varyansı verilmiştir. Bu desen, puanlayıcıların her birinin birbirleriyle eşit sayıda öğrenciyi puanlaması nedeniyle dengelenmiş desen örneğidir.

Her bir puanlayıcının birbirleriyle eşit sayıda öğrenciyi puanlamasıyla oluşturulmuş dengelenmiş (ö:p) x g deseni için Eşitlik 5'te yer alan her bir varyans bileşeni varyans analizi kullanılarak Çizelge 1'deki eşitlikler yardımıyla kestirilir.

Çizelge 1. Dengelenmiş Desen İçin Varyans Bileşenlerinin Kestirilmesi

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Kestirilen Varyans Bileşeni
g	$SS_{(g)}$	n_g-1	$MS_{(g)}$	$\sigma^2(g)=[MS_{(g)}-MS_{(gp)}]/n_\delta n_p$
p	$SS_{(p)}$	n_p-1	$MS_{(p)}$	$\sigma^2(p)=[MS_{(p)}-MS_{(\delta:p)}-MS_{(gp)}+MS_{(g\delta:p)}]/n_g n_\delta$
ö:p	$SS_{(\delta:p)}$	$n_p(n_\delta-1)$	$MS_{(\delta:p)}$	$\sigma^2(\delta:p)=[MS_{(\delta:p)}-MS_{(g\delta:p)}]/n_g$
gp	$SS_{(gp)}$	$(n_g-1)(n_p-1)$	$MS_{(gp)}$	$\sigma^2(gp)=[MS_{(gp)}-MS_{(g\delta:p)}]/n_\delta$
gö:p	$SS_{(g\delta:p)}$	$n_p(n_\delta-1)(n_g-1)$	$MS_{(g\delta:p,e)}$	$\sigma^2(g\delta:p,e)=MS_{(g\delta:p)}$

Çizelge 1’de elde edilen varyans bileşenleri yardımıyla da G ve Phi katsayıları elde edilir. Aynı görevler doğrultusunda puanlayıcıların her birinin puanladığı öğrenci sayısının birbirleriyle eşit olduğu, görevlerin ölçme konusu ve tüm değişkenlerin tesadüfi olarak belirlendiği dengelenmiş (ö:p) x g deseni için hesaplanan genellenebilirlik ve Phi katsayıları Eşitlik 9 ve Eşitlik 11’de verilmiştir.

Buraya kadar olan kısımda dengelenmiş desen ele alınmıştır. Fakat pratikte dengelenmemiş desenlere ait durumlarla daha sık karşılaşılmaktadır. Çünkü yukarıdaki örnekte verildiği 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 (balanced) desenlerde, değişkenin her bir düzeyi için gözlemlerin sayısının aynı sayıda olduğu belirtilmişti. Fakat her durumda değişken başına elde edilen gözlemler eşit sayıda olmayabilir. Kayıp (missing) veri ya da değişkenin düzeylerine ait gözlem sayısının değiştiği durumlarda dengelenmemiş desenler ortaya çıkmaktadır (Brennan, 2001).

Bir deęişkenin koşullarına ait hücrelerdeki gözlem sayısının eşit olmadığı durumlarda veri dengelenmemiş olarak adlandırılır (Kaufman ve Schering, 2007). Dengeli verilerin analizi nispeten kolay ve daha yaygın olmasına rağmen dengelenmemiş veriler ile varyans bileşenlerinin kestirimi ise geniş istatistiksel işlemler gerektirmektedir (Jarjoura ve Brennan, 1981).

Deęişkenin 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 atmamak yerine dengelenmemiş desenlerle varyans bileşenlerini hesaplamak gerekir (Bell, 1985). Büyük veri gruplarında veri indirgemesi yapmadan dengelenmemiş desenlerle de sonuç elde etmek mümkündür.

Çizelge 1'de dengelenmiş (ö:p) x g deseni örneğine ait varyans bileşenleri kestirilmiştir. Aynı örnekten hareketle bu kez puanlayıcıların dönüşümlü olarak birbirlerinden farklı sayıda öğrenciyi puanladığı dengelenmemiş (ö:p) x g (ö: öğrenci, p: puanlayıcı ve g: görev olmak üzere) desenine ait varyans bileşenleri ise Çizelge 2'deki eşitlikler yardımıyla kestirilmektedir.

Çizelge 2'de varyans bileşenlerini kestirmede kullanılan $r_{\ddot{o}}$ ve $t_{\ddot{o}}$ ise Eşitlik 14'teki gibidir.

n_+ = Desendeki toplam gözlem sayısı

$n_{\ddot{o}+} = \sum_p n_{\ddot{o}:p}$ = Toplam öğrenci sayısı

$$r_{\ddot{o}} = \sum_p \frac{n_{\ddot{o}:p}^2}{n_{\ddot{o}+}} \quad t_{\ddot{o}} = \frac{n_{\ddot{o}+} - r_{\ddot{o}}}{n_p - 1} \quad (14)$$

**Çizelge 2. Dengelenmemiş Desen İçin Varyans Bileşenlerinin
Kestirilmesi**

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Kestirilen Varyans Bileşeni
g	$SS_{(g)}$	n_g-1	$MS_{(g)}$	$\sigma^2(g)=[MS_{(g)}-r_{\delta}MS_{(gp)}/t_{\delta}+(r_{\delta}-t_{\delta})MS_{(g\delta:p)}/t_{\delta}]/n_{\delta+}$
p	$SS_{(p)}$	n_p-1	$MS_{(p)}$	$\sigma^2(p)=[MS_{(p)}-MS_{(\delta:p)}-MS_{(gp)}+MS_{(g\delta:p)}]/n_g t_{\delta}$
ö:p	$SS_{(\delta:p)}$	$n_{\delta+} - n_p$	$MS_{(\delta:p)}$	$\sigma^2(\delta:p)=[MS_{(\delta:p)}-MS_{(g\delta:p)}]/n_g$
gp	$SS_{(gp)}$	$(n_g-1)(n_p-1)$	$MS_{(gp)}$	$\sigma^2(gp)=[MS_{(gp)}-MS_{(g\delta:p)}]/t_{\delta}$
gö:p	$SS_{(g\delta:p)}$	$(n_g-1)(n_{\delta+}-n_p)$	$MS_{(g\delta:p,e)}$	$\sigma^2(g\delta:p,e)=MS_{(g\delta:p)}$

Dengelenmemiş (ö:p) x g deseni için kestirilen bağıl hata varyansı, dengelenmiş desen ile kestirilen bağıl hata varyansından biraz farklılık göstermektedir. Böylece puanlayıcıların her birinin puanladığı öğrenci sayısının birbirlerinden farklı olduğu, görevlerin ölçme konusu ve tüm değişkenlerin tesadüfi olarak belirlendiği dengelenmemiş (ö:p) x g deseni için kestirilen genellenebilirlik katsayısı Eşitlik 15'teki gibi olmaktadır.

$$E\rho^2 = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\delta)} \quad (15)$$

$$\sigma^2(\tau) = \sigma^2(g)$$

$$\sigma^2(\delta) = \frac{\sigma^2_{(gp)}}{\tilde{n}_p} + \frac{\sigma^2_{(g\delta:p)}}{n_{\delta+}}$$

$$\tilde{n}_p = \frac{n_{\delta+}^2}{\sum_p n_{\delta:p}^2}$$

Aynı durum için Phi katsayısı ise;

$$\Phi = \frac{\sigma^2(\tau)}{\sigma^2(\tau) + \sigma^2(\Delta)} \quad (16)$$

$$\sigma^2(\tau) = \sigma^2(g)$$

$$\sigma^2(\Delta) = \frac{\sigma^2(p)}{\check{n}_p} + \frac{\sigma^2(\ddot{o}: p)}{n_{\ddot{o}+}} + \frac{\sigma^2(gp)}{\check{n}_p} + \frac{\sigma^2(g\ddot{o}: p)}{n_{\ddot{o}+}}$$

Eşitlik 16'daki gibi olmaktadır.

Çizelge 1 ve 2, Eşitlik 9, 11, 15 ve 16'dan da görülmektedir ki; dengelenmiş ve dengelenmemiş desenlere göre her bir varyans bileşeninin kestirilmesinde kullanılan eşitlikler ile G ve Phi katsayılarının hesaplanmasında kullanılan formüller değişmektedir. Bununla birlikte dengelenmemiş desenlerde ölçme konusunun değiştirilmesine bağlı olarak da eşitliklerde kullanılan bazı formüller değişmekte, daha öncede belirtildiği gibi geniş istatistiksel işlemler gerekmektedir.

Literatürdeki genellenebilirlik kuramıyla ilgili yapılan çalışmalar incelendiğinde, yurt dışında yapılan çalışmaların büyük çoğunluğunda dengelenmiş veri yapılarının kullanılmasının yanında, dengelenmemiş verilerle her iki veri yapısının bir arada kullanıldığı çalışmaların olduğu görülmektedir (Bergus ve Kreiter, 2007; Clauser, Swanson ve Clyman, 1996; Dory, Gagnon ve Charlin, 2010; Jeon, Lee, Hwang ve Kang, 2009; Lee ve Frisbie 1999; Nie, Yeo ve Lau, 2007; Ødegård, Hagtvet ve Bjørkly, 2008; Sharma ve Weathers, 2003; Wei and Haertel, 2011; Yin, 2005; Zibrowski, Myers, Norman, Goldszmidt, 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, 2009; Güler, ve Gelbal, 2010; Güler ve Çetin, 2010; Güler, 2011;

Nalbantođlu, 2009; Taşdelen, 2009; Yelbođa, ve Tavşancıl, 2010; Yılmaz Nalbantođlu ve Gelbal, 2011). Ayrıca yapılan bu çalışmaların tümünde dengeli veri tasarımlarının kullanıldığı dikkati çekmektedir. Bununla birlikte dengelenmemiş veri tasarımlarının 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 puanlayıcı ve öğrenci değişkenlerinin yuvalandığı ve puanlayıcıların her birinin puanladığı öğrenci sayısı bakımından aralarında farklılıklar olduğu dengelenmemiş veri yapılarının güvenilirlik üzerindeki etkisinin araştırılması gerekmektedir. Böylece Türkiye’de genellenabilirlik kuramıyla yapılan 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. Bu doğrultuda dengelenmiş ve dengelenmemiş veri yapılarıyla oluşturulan desenlerden elde edilen güvenilirliğin nasıl olacağı, aynı verilerle dengelenmemiş desenlerle hesaplanan güvenilirlik katsayılarının dengeli yapılara göre güvenilirliği etkileyip etkilemeyeceği ve desenler için kestirilen varyans bileşenleri arasında farklılık olup olmayacağı belirlenmesi gereklilik olarak görülmüştür.

Amaç

Eđitimin her kademesinde, özellikle de bir meslek grubuna ait becerilerin kazandırıldığı yükseköđretim programlarında öğrencilerin gösterdiği becerilerle ilgili karar alırken kullanılan puanların güvenilirliğinin belirlenmesi önem taşımaktadır. Bu doğrultuda araştırmanın genel amacı, tıp eğitiminde öğrencilerin mesleki becerilerinin değerlendirildiği OSCE sınavı İntramuskuler Enjeksiyon yapma istasyonu verileriyle, puanlanan toplam 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. Öğrenci (ö) ve puanlayıcı (p) değişkenlerinin yuvalanmış, görev (g) değişkeninin ise çapraz tasarlandığı dengelenmiş (ö:p) x g deseninin genellenebilirlik kuramı sonuçları nasıldır?
 - 1.1. Dengelenmiş desende G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları nasıldır?
 - 1.2. Dengelenmiş (ö:p) x g deseninde puanlayıcı ve öğrenci sayılarının artırılıp azaltılması senaryolarına göre karar çalışmasıyla kestirilen G ve Phi katsayıları sonuçları nasıldır?
2. Öğrenci (ö) ve puanlayıcı (p) değişkenlerinin yuvalanmış, görev (g) değişkeninin ise çapraz tasarlandığı dengelenmemiş (ö:p) x g deseninin genellenebilirlik kuramı sonuçları nasıldır?
 - 2.1. Dengelenmemiş desende G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları nasıldır?
 - 2.2. Dengelenmemiş (ö:p) x g deseninde puanlayıcı ve öğrenci sayılarının artırılıp azaltılması senaryolarına göre karar çalışmasıyla kestirilen G ve Phi katsayıları sonuçları nasıldır?
3. Dengelenmiş ve dengelenmemiş (ö:p) x g desenlerinden elde edilen genellenebilirlik kuramı sonuçları değişmekte midir?
 - 3.1. Her iki desenden elde edilen varyans bileşenleri ve toplam varyansı açıklama oranları değişmekte midir?

- 3.2. Her iki desende puanlayıcı ve öğrenci sayılarının artırılıp azaltılmasıyla senaryolara göre yapılan karar çalışmalarından kestirilen G ve Phi katsayıları değişmekte midir?

Önem

Üst düzey zihinsel becerilerin ölçüldüğü performans sınavları, eğitim öğretimin her kademesinde önem taşımaktadır. Özellikle bir meslek alanına ilişkin becerilerin kazandırıldığı yükseköğretim programlarında kazandırılan becerilerin ne düzeyde olduğunu belirlemede ve bu doğrultuda öğrenci hakkında alınacak kararların güvenilirliği ve geçerliği açısından performans sınavlarının önemi daha da büyüktür.

Performans değerlendirme çalışmaları puanlayıcının aşırı katı ya da cömert olması, puanlayıcılar arası tutarsızlık, görev zorluğu gibi birçok hata kaynağından etkilenmektedir. Bu nedenle elde edilen performans puanlarını kullanmadan önce performans değerlendirme çalışmalarını etkileyen hata kaynaklarını tek tek ve birbirleriyle etkileşimlerini de dikkate alan genellenebilirlik kuramı ile güvenilirlik incelemeleri yapmak büyük önem taşımaktadır.

Araştırma verilerinin kullanıldığı OSCE sınavı birden fazla istasyondan oluşmaktadır. Sınava katılan öğrenci sayısının çok olması nedeniyle öğrencilerin istasyonlara ait performanslarını değerlendirmek uzun zaman almaktadır. Bu sebeple sınavda 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 pratikte kullanabilecekleri uygulama ve farklı puanlama durumlarına göre güvenilirliğin ve puanlayıcı etkisinin araştırılması bakımından da katkı sağlayacağı düşünülmektedir.

Türkiye’de performans değerlendirme çalışmalarında genellenebilirlik kuramı ile yapılan araştırmalar sınırlı kalmakla birlikte, yapılan araştırmalar incelendiğinde dengelenmiş veri yapılarıyla oluşturulan desenlerin kullanıldığı görülmektedir. Bu açıdan, araştırmanın gerçek durum verilerine daha uygun olan dengelenmemiş veri yapılarıyla oluşturulan desenler ile genellenebilirlik kuramı analizleri yapılması ve puanlayıcıların dönüşümlü olarak aynı ya da farklı sayıda öğrenciyi puanlamasıyla oluşturulmuş dengelenmiş ve dengelenmemiş desenlerle elde edilen sonuçların karşılaştırılması bakımından da kuramsal bir katkı sağlayacağı düşünülmektedir. Ayrıca dengelenmemiş veriler üzerinde genellenebilirlik kuramı analizleriyle güvenilirlik incelemesi yapılmasının araştırmacıların veri yapılarına uygun sonuçlar elde etmeleri konusunda önemli bilgiler vereceği de düşünülmektedir.

Sayıtlar

OSCE sınavı İntramuskuler Enjeksiyon Yapma istasyonuna katılan öğrenciler ilgili istasyon bakımından gerçek becerilerini yansıtmışlardır.

Sınırlılıklar

Bu araştırma, 2010-2011 eğitim-öğretim yılı Hacettepe Üniversitesi Tıp Fakültesi’nde uygulanan OSCE sınavı İntramuskuler Enjeksiyon Yapma becerilerinin ölçüldüğü klinik sınava katılan 1. sınıf öğrencilerinden değerlendirme formundaki verilerin tam olarak doldurulduğu öğrencilerle sınırlıdır.

Tanımlar

Genellenebilirlik Çalışması (G Çalışması): Farklı koşullardaki ölçmelerden kaynaklanan çeşitli değişkenlik kaynaklarının ölçme hatalarını araştırmak ve değişkenlik kaynaklarına ait varyans bileşenlerini hesaplamak için yapılan çalışmadır.

Değişkenlik Kaynağı (Facet): Potansiyel hata kaynağı olarak düşünülen zaman, puanlayıcı, görev, madde gibi etkilerdir.

Koşul (Condition): Bir değişkenlik kaynağının düzeyi olarak tanımlanmaktadır. Örneğin; görev değişkenlik kaynağı olarak ele alındığında çalışma kapsamındaki her bir görev değişkenlik kaynağının koşuludur.

Çapraz (Crossed): Bir değişkenlik kaynağının bütün koşullarının diğer değişkenlik kaynağının bütün koşullarıyla örtüştüğü duruma denir.

Yuvalanmış (Nested): Bir değişkenlik kaynağının koşullarının diğer değişkenlik kaynağının bazı koşullarıyla örtüştüğü duruma denir.

Dengeli Tasarım (Balanced Design): Gözlem sayısının değişkenin tüm koşullarında eşit ve dengeli dağılmasıdır.

Dengelenmemiş Tasarım (Unbalanced Design): Gözlem sayısının değişkenin tüm koşullarında eşit dağılmamasıdır.

Kısaltmalar

G Kuramı : Genellenebilirlik Kuramı

G Çalışması : Genellenebilirlik Çalışması

K Çalışması : Karar Çalışması

G Katsayısı : Genellenebilirlik Katsayısı

Phi Katsayısı : Güvenirlilik Katsayısı

$\sigma^2(T)$: Evren Puanı Varyansı

$\sigma^2(\delta)$: Bağıl Hata Varyansı

$\sigma^2(\Delta)$: Mutlak Hata Varyansı

E_p^2 : Genellenebilirlik Katsayısı

Φ : Güvenirlik Katsayısı

x : Çaprazlama

“:” : Yuvalama

\ddot{o} : Öğrenci

p : Puanlayıcı

g : Görev

$g \times p$: Görev x puanlayıcı ortak etkisi

$\ddot{o} : p$: Öğrenci : puanlayıcı ana etkisi

$g\ddot{o} : p$: Görev x öğrenci: puanlayıcı artık etkisi

BÖLÜM II

İLGİLİ ARAŞTIRMALAR

Bu bölümde, Türkiye’de yapılan Genellenebilirlik Kuramı çalışmaları ile yurt dışında yapılan dengelenmiş (balanced) ve/veya dengelenmemiş (unbalanced) desenlerle Genellenebilirlik Kuramı’nın kullanıldığı çalışmalar verilmiştir.

Yurt İçinde Yapılan Çalışmalar

Atılğan (2004), “Genellenebilirlik Kuramı ve Çok Değişken Kaynaklı Rasch Modelinin Karşılaştırılması” adlı çalışmasında iki yıl ardına yapılan müzik öğretmenliği özel yetenek seçme sınavı verilerini kullanarak Genellenebilirlik Kuramı ve Çok Değişken Kaynaklı Rasch Modeli analizleri yapmıştır. Araştırmanın son kısmında ise Genellenebilirlik Kuramı ve Çok Değişken Kaynaklı Rasch Modeli istatistikleri karşılaştırılmıştır. Araştırma sonuçlarına göre, Genellenebilirlik Kuramı tek değişkenli ve çok değişkenli modellerinin aynı ölçme durumu için alt boyutlardan oluşan testlerde farklı sonuçlar ürettiği, Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch yaklaşımlarıyla elde edilen sonuçların ise kısmen tutarlı olduğu bulunmuştur.

Yelboğa (2007), “Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre Güvenirliğin Bir İş Performansı Ölçeği Üzerinde İncelenmesi” adlı çalışmasında 2005-2006 yılları arasında İş performansı ölçeği kullanarak Klasik Test Kuramı ve Genellenebilirlik Kuramından elde edilen verilerle güvenilirlik katsayılarını karşılaştırmıştır. Çalışmada güvenilirlik katsayılarının karşılaştırılmasında Klasik test kuramında, test tekrar test ve Cronbach alfa güvenilirlik katsayıları, değerlendiriciler arası güvenilirlik için Kendall’ın Konkordans katsayısı, Genellenebilirlik Kuramı’nda ise çok değişkenli modelle

G ve Phi katsayıları hesaplanmıştır. Araştırmadan elde edilen sonuçlara göre, iş performansı ölçeğinde Klasik Test Kuramı ve Genellenebilirlik Kuramı'nın çok değişkenli modeline göre elde edilen güvenilirlik katsayılarının uyumlu sonuçlar verdiği bulunmuştur.

Güler (2008), Klasik Test Kuramı, Genellenebilirlik Kuramı ve Rasch Modeli üzerine yaptığı çalışmada matematik başarısının ölçülmesinden elde edilen puanların güvenilirliklerini hesaplamış ve üç kuramdan elde edilen sonuçları karşılaştırmıştır. Puanların güvenirligi Klasik Test Kuramı'nda Cronbach Alfa, puanlayıcılar arası uyumun belirlenmesinde Kendall Konkordans katsayısı ve puanlayıcılar arası korelasyon katsayısı kullanılmıştır. Genellenebilirlik kuramı analizlerinde tümüyle çaprazlanmış bxgxp deseni kullanılarak genellenebilirlik ve güvenilirlik katsayıları hesaplanmış, çok değişkenli Rasch modeli ile de birey, puanlayıcı ve madde boyutlarına ait güvenilirlik hesaplanmıştır. Araştırma sonuçlarına göre, matematik başarısını ölçmek için kullanılan ölçme aracının güvenilir sonuçlar verdiği, puanlayıcıların birbirleriyle uyumlu puanlamalar yaptığı belirlenmiştir. Ayrıca üç kurama göre yapılan analizlerin öğrenci boyutunda güvenilirliklerin birbirine yakın sonuçlar verdiği görülmüştür.

Yelboğa (2008), "Genellenebilirlik Kuramı İle Bir Personel Seçme Yönteminde Güvenirliğin Değerlendirilmesi" adlı çalışmasında otomotiv firmasına personel seçme sürecinde yapılandırılmış mülakat formu kullanarak elde edilen puanların güvenirliginin değerlendirilmesinde G ve Phi katsayılarını hesaplamış ve alternatif karar çalışmaları yapmıştır. Araştırmadan elde edilen bulgular doğrultusunda görevlerin güçlük bakımından birbirine yakın olduğu, puanlayıcıların birbirlerinden oldukça farklı değerlendirme yaptıkları sonuçlarına ulaşılmıştır. Ayrıca yapılan karar çalışmasında G ve Phi katsayılarını artırmak için puanlayıcı sayısının artırılmasından ziyade madde sayısının artırılması gerektiği tespit edilmiştir.

Deliceoğlu (2009), "Futbol Yetilerine İlişkin Dereceleme Ölçeğinin Genellenebilirlik ve Klasik Test Kuramına Dayalı Güvenirliklerinin Karşılaştırılması" adlı çalışmasında futbolcuların teknik yetilerinin tespit

edilmesinde kullanılan ölçeğin Klasik Test Kuramı ve Genellenebilirlik Kuramı'na dayalı olarak güvenilirlik düzeyi saptanmış ve sonuçlar karşılaştırılmıştır. Çalışmada Genellenebilirlik Kuramı'na göre G ve Phi katsayıları hesaplanmış, Klasik Test Kuramı'na göre ise Cronbach Alfa ve Kendall uyuşum katsayısı hesaplanmıştır. Araştırma sonuçlarına göre elde edilen güvenilirlik katsayıları iç ölçütlere göre incelendiğinde G katsayısı ve Cronbach Alfa katsayıları beklenen değerlerinden yüksek olduğu, Phi katsayısı ile Kendall W güvenilirlik katsayılarının beklenen değerlerinden düşük olduğu tespit edilmiştir. Ayrıca potansiyel hata kaynaklarının fazla olduğu durumlarda G Kuramına dayalı güvenilirlik tahmininin, Klasik Test Kuramı'na dayalı güvenilirlik kestirimlerine alternatif olabileceği belirtilmiştir.

Nalbantoğlu (2009), "Performans Ölçümlerinde Genellenebilirlik Kuramıyla Farklı Desenlerin Karşılaştırılması" adlı çalışmasında bir performans puanlama örneğinde öğrencilerin birden fazla puanlayıcı tarafından birlikte ve dönüşümlü olarak puanlanmasıyla oluşturulan desenlerle yapılan G ve K çalışmaları sonuçları karşılaştırılmıştır. Araştırmada puanlayıcıların öğrencileri aynı iletişim becerileri değerlendirme formuyla 15 görev doğrultusunda birlikte ve dönüşümlü olarak puanlanmasıyla oluşturulan $\bar{o} \times g \times p$ ve $(\bar{o}: p) \times g$ desenleri (\bar{o} : öğrenci, g : görev, p : puanlayıcı) için ayrı ayrı G ve K çalışması yapılmıştır. Her iki desenden elde edilen sonuçlar karşılaştırılmıştır. Araştırmadan elde edilen sonuçlara göre, her iki desende değişkenler için kestirilen varyans değerlerinin birbirleriyle paralellik gösterdiği, sınavda puanlayıcıların öğrencileri puanlamada aralarında farklılık olmadığı bulunmuştur. Ayrıca yapılan karar çalışmaları sonucunda her iki desendeki G ve Phi katsayıları arasında çok fark olmamakla birlikte $(\bar{o}:p) \times g$ deseninde katsayıların daha yüksek çıkma eğiliminde olduğu tespit edilmiştir. Elde edilen sonuçlardan hareketle çok sayıda öğrencinin bulunduğu performans sınavlarında puanlayıcılar arası tutarlılık sağlandığında öğrencilerin puanlayıcıların hepsi tarafından tek tek puanlanması yerine puanlayıcıların belli sayıdaki öğrencileri dönüşümlü puanlamasının zaman, iş gücü ve ekonomiklik açısından daha uygun olabileceği sonucuna varılmıştır.

Taşdelen (2009), Nedelsky ve Angoff standart belirleme yöntemleri ile elde edilen kesme puanlarını Genellenebilirlik Kuramı kullanılarak karşılaştırmıştır. Çalışmada kullanılan standart belirleme yöntemlerinin uygulanması için gereken en uygun puanlayıcı sayısı belirlenmeye çalışılmıştır. Araştırmada, 2008 yılı ilköğretim 6. sınıf Seviye Belirleme Sınavı'nda yer alan ve 16 maddeden oluşan fen ve teknoloji sorularının Ankara ilinde görev yapan 40 fen ve teknoloji uzmanı tarafından adı geçen standart belirleme yöntemlerine uygun şekilde puanlanması sonucunda elde edilen veriler kullanılmıştır. Araştırma bulgularına göre G çalışmasında her iki yöntemde de, maddeler arası farklılıkların ortaya çıkarılabildiği ve puanlayıcılar arasında bir tutarlılığın olduğu tespit edilmiştir. Ayrıca yapılan karar çalışması ile Angoff ve Nedelsky yöntemleri ile standart belirlemede 10 puanlayıcının yeterli olacağı sonucuna varılmıştır.

Güler ve Çetin (2010), "Genellenebilirlik Kuramı ve Klasik Test Kuramı Açısından Gözetmen Sayısı Etkisinin İncelenmesi" adlı çalışmalarında çoktan seçmeli test uygulamalarında farklı sayıda gözetmen bulundurulmasına bağlı olarak değişen durumlarda öğrencilerden elde edilen puanların güvenilirliği Cronbach alfa ve genellenebilirlik katsayısı ile incelenmiştir. Genellenebilirlik kuramı ile hesaplanan genellenebilirlik katsayısı tek değişkenlik ve iki değişkenlik kaynaklı olarak iki şekilde hesaplanmıştır. Araştırma bulgularına göre tek değişkenlik kaynaklı çalışmada Cronbach alfa ve G katsayıları birbirine eşit çıkmıştır.

Güler (2011), "Rastgele Veriler Üzerinde Genellenebilirlik Kuramı ve Klasik Test Kuramı'na Göre Güvenirliğin Karşılaştırılması" adlı çalışmasında ölçme hatası yüksek, güvenilirliği düşük olan veriler üzerinde iki kurama dayalı olarak güvenirlilik hesaplamış ve elde edilen sonuçları karşılaştırmıştır. Çalışmada Klasik Test Kuramı'na göre güvenirliliğin hesaplanmasında Cronbach Alfa katsayısı kullanılmıştır. Araştırma bulguları doğrultusunda ise tek değişkenlik kaynaklı tümüyle çapraz desen için kestirilen genellenebilirlik katsayısı değerleri ile Cronbach Alfa değerlerinin birbirine eşit çıktığı sonucuna varılmıştır.

Yurt içinde yapılan çalışmalara bakıldığında, genellenebilirlik kuramı sonuçlarının klasik test kuramı ve/veya çok değişken kaynaklı Rasch modeli ile karşılaştırıldığı çalışmaların ağırlık kazandığı görülmektedir. Bununla birlikte, yapılan çalışmalarda değişkenlik kaynaklarının çapraz tasarlanması sonucu oluşturulmuş desenlerin sıklıkla kullanıldığı ve genellenebilirlik kuramı analizleri için dengelenmiş desenlerin kullanıldığı görülmektedir. Ayrıca genellenebilirlik kuramı ile yapılan çalışmaların Türkiye’de son yıllarda ağırlık kazanmasıyla birlikte yurt dışındaki yayınlara göre hala sınırlı sayıda olduğu da dikkati çekmektedir.

Yurt Dışında Yapılan Çalışmalar

Bu bölümde, dengelenmiş (balanced) ve/veya dengelenmemiş (unbalanced) desenlerle Genellenebilirlik Kuramı’nın kullanıldığı araştırma bulguları özetlenmiştir.

Bell (1985), genellenebilirlik kuramı yazılım sorunu adlı çalışmasında dengelenmiş (balanced) ve dengelenmemiş (unbalanced) desenlere ait örnek kestirimler yapmıştır. Bu örnek kestirimler için öğrenci (ö), okul (o) ve soru (s) ile gösterilmek üzere farklı okullardan gelen öğrencilerin kullanıldığı (ö:o) x s deseni ile dengelenmiş ve dengelenmemiş desenleri kullanmıştır. Çalışmada MIVQUE metodunun bazı sınırlılıkları ortadan kaldırdığını ve dengelenmiş desenler için SAS veya GENOVA programlarını kullanmanın daha uygun olduğu, dengelenmemiş desenler için varyans kestirimleri yapmanın gerekli olduğu ve dengelenmemiş desenlerle analiz işlemlerinin teorik ve pratik zorluğundan ve yazılımların eksikliğinden dolayı dengeli veri yapıları oluşturulabildiği sürece dengelenmiş desenler kullanmayı önermiştir.

Lee ve Frisbie (1999), yaptıkları araştırmada ulusal standart test uygulaması olan Iowa Temel Beceriler Testi (ITBS) ve Iowa Öğretim Geliştirme Testi (ITED) verilerini kullanarak genellenebilirlik kuramıyla güvenilirlik kestirimleri yapmışlardır. Araştırmada testlerde her bir pasaja ait madde sayısı farklı olduğunda genellenebilirlik kuramı analizleri için

dengelenmemiş b × (m:p) deseni (b:birey, m: madde ve p: pasaj olmak üzere) kullanılmıştır. Dengelenmemiş b x (m:p) deseni ile yapılan analizlere göre ise güvenilirliği artırmak için madde sayısını artırmak yerine pasaj sayısını artırmanın daha etkili bir yol olduğu sonucuna varılmıştır.

Lee ve Fitzpatrick (2003), yaptıkları çalışmada genellenebilirlik kuramı çerçevesinde karar çalışmasıyla üç farklı grup için okul değişkeni için standart hataları karşılaştırmayı amaçlamıştır. Çalışmada öncelikle genellenebilirlik kuramıyla ö:(oxf) desenine (ö: öğrenci, o:okul ve f: test formu olmak üzere) ait varyans bileşenlerini kestirilmiş, daha sonra ise karar çalışması yapılmıştır. Desende her bir okulda testi alan öğrenci sayısı farklılık gösterdiğinden dengelenmemiş desen kullanılmıştır. Araştırma sonuçlarına göre test formu etkisinden gelen farklılıkların olmadığı, artık varyans etkisinin yüksek olduğu, okul standart hatalarının öğrenci sayısına bağlı olarak değiştiği, toplam öğrenci sayısı küçük olduğunda üç farklı grupta standart hataların farklılık gösterdiği tespit edilmiştir.

Sharma ve Weathers (2003), yaptıkları çalışmada genellenebilirlik kuramı kullanarak uluslar arası araştırmalarda kullanılan tüketici etnik merkezilik ölçeğinin madde sayısı hakkında karar çalışması yapmayı amaçlamışlardır. Ayrıca araştırmacılar ülkeler arası ölçme değişmezliğini belirlemede genellenebilirlik kuramı ve doğrulayıcı faktör analizini kullanarak elde edilen sonuçları da karşılaştırmıştır. Araştırmada dört ülkeye ait veriler kullanılmış olup her ülkeden farklı sayıda bireye ait ölçümler olduğundan dengelenmemiş desen oluşturulmuştur. Fakat dengelenmemiş desene ilişkin varyans kestirimleri karmaşık olduğundan çalışma üç yaklaşımla yürütülmüştür. Bu yaklaşımlar her bir ülkeden eşit sayıda birey kullanılarak dengelenmiş, her ülkeden farklı sayıda birey kullanılarak dengelenmemiş desen yaratan tüm veriler ve varyans kestirimlerinin sağlamlığı için farklı örneklem büyüklükleri kullanılan dengelenmiş verilerdir. Her bir yaklaşımla kestirilen varyans bileşenleri çok benzer bulunduğu için araştırmada dengelenmiş örneğe ilişkin bulgular sunulmuştur. Araştırma sonuçlarına göre, kabul edilebilir genellenebilirlik düzeyinde ölçeğin madde sayısının 11 ile 17 arasında seçilebileceği tespit edilmiştir. Ayrıca faktör analizinin ölçeğin

kapsamı hakkında ve ölçek geliştirme çalışmalarında genellenebilirlik kuramına göre daha açıklayıcı bilgiler verdiği tespit edilmiştir. Bununla birlikte genellenebilirlik kuramının da karar çalışması ile gelecek çalışmalar için ön tahminde bulunulabilmesi açısından önemli olduğu belirtilmiştir. Ayrıca araştırmacılar iki kuramın da sınırlılıklarının yanında faydalarının bulunduğunu, iki kuramın da birbirini tamamlayıcı olduğunu da belirtmişlerdir.

Barneveld (2005), öğrencilerin performans puanlarının güvenilirliğini belirlemek amacıyla yaptığı çalışmada, öğrencilerin puanlayıcılarla yuvalandığı dengelenmemiş (ö:p) x m deseni (ö: öğrenci, p:puanlayıcı ve m:madde olmak üzere) genellenebilirlik analizleri yapılmıştır. Çalışmada genellenebilirlik kuramı analizleri için urGENOVA programını kullanmıştır. Araştırma sonuçlarına göre öğrenci performanslarının bir puanlayıcıdan diğerine farklılık gösterdiği, öğrenci performansını değerlendirmede güvenirliliğin düşük çıktığı belirlenmiştir. Elde edilen güvenirliliğin ise puanlayıcı sayısı artırıldığında yükseldiği tespit edilmiştir.

Yin (2005), çalışmasında çok değişkenli genellenebilirlik kuramı analizleriyle farklı alt testlerden oluşan çok aşamalı avukatlık sınavına ait G ve K çalışması ile sınavın yapısını incelemiştir. Analizlerde her bir alt testte farklı maddelerin bulunduğu ve bireylerin tüm maddeleri cevapladığı çok değişkenli dengelenmemiş $b \times m^o$ deseni (b:birey, m:madde olmak üzere) kullanılmıştır. Testte yer alan her bir alt testteki madde sayısı farklı olduğu için desen dengelenmemiştir. Çalışmada sınavın uygulandığı son üç yıla ait veriler kullanılmış ve her bir yıla ait veriler genellenebilirlik kuramıyla ayrı ayrı incelenmiştir. Araştırma bulgularına göre temmuz ve şubat aylarında yapılan sınavların varyans ve kovaryans kestirimleri ile genellenebilirlik katsayıları hemen hemen aynı çıkmıştır. Ayrıca elde edilen G katsayılarının (0,86 ile 0,90 aralığında) yüksek çıktığı tespit edilmiştir.

Bergus ve Kreiter (2007) yaptıkları çalışmada farklı yıllara ait klinik sınavların güvenirliliğini araştırmışlardır. Çalışmada tıp fakültesi öğrencilerinin psikiyatri, jinekoloji, dâhiliye, cerrahi ve pediatri gibi beş staja ait ve her bir stajda iki ile dört arasında değişen olgunun incelendiği veriler kullanılmıştır.

Öğrencilerin tüm stajlara girdiği fakat yıllara göre her bir stajdaki olguların değiştiği dengelenmemiş $o \times (o:s)$ desenine (ö: öğrenci, s: staj ve o:olgu olmak üzere) ait varyans bileşenleri kestirilmiş ve karar çalışması yapılmıştır. Çalışma bulgularından hareketle stajlara ait puanların orta derecede güvenilir olduğu sonucuna ulaşılmıştır.

Pedersen, Hagtvet, Karterud (2007), psikiyatrik semptomların ve sosyal, mesleki düzeyler ile ilgili genel işlevselliğin değerlendirilmesi (Global Assessment of Functioning-GAF) ölçeğinin güvenilirliğinin araştırılmasında genellenebilirlik kuramını kullanmıştır. Çalışmada GAF puanlarındaki hata varyansları araştırılmış ve farklı senaryo durumlarına göre karar çalışmasıyla güvenilirlik araştırılmıştır. Analizlerde her bir birimdeki puanlayıcı sayısı farklı olduğundan dengelenmemiş $b \times (p:k)$ deseni (b: birey, p: puanlayıcı ve k: konu olmak üzere) kullanılmıştır. Sonuç olarak, ölçek puanlarının kabul edilebilir derecede güvenilir olduğu, konu başına iki puanlayıcının kullanılmasının güvenilirlik açısından daha etkili olduğu tespit edilmiştir.

Malhotra ve Sharma (2008), yaptıkları çalışmada ölçme eşdeğerliğini belirlemede doğrulayıcı faktör analizinin yanında genellenebilirlik kuramını da kullanmışlardır. Çalışmada genellenebilirlik kuramı analizlerinde her bir grupta eşit ve farklı sayıda öğrencilerin bulunduğu ve tüm öğrencilerin aynı maddeleri aldığı dengelenmiş ve dengelenmemiş $(ö:g) \times m$ (ö: öğrenci, g: grup ve m:madde olmak üzere) deseni kullanmışlardır. Çalışmanın genellenebilirlik kuramı boyutunda yapılan analizlerde dengelenmiş ve dengelenmemiş desenlerle elde edilen analiz sonuçlarında her iki desenle alt boyutlara göre hesaplanan güvenilirlik katsayıları arasında büyük farklılık olmadığı tespit edilmiştir. Ayrıca çalışma sonuçlarına göre küçük örneklem büyüklüklerinde doğrulayıcı faktör analizi yerine genellenebilirlik kuramının kullanılabileceğini önermişlerdir.

Ødegård, Hagtvet ve Bjørkly (2008) tarafından yapılan çalışmada bireylerin kendi meslekleri ile ilgili meslekler arası işbirliğine dayalı algılarını ölçmek için hazırlanan çok düzeyli anketin (questionnaires) farklı kaynaklardan gelen varyans bileşenlerini kestirmek için genellenebilirlik

kuramı kullanılmıştır. Çalışma kapsamında farklı meslek grubundan 157 kişiye uygulanan çok düzeyli anket (b:g)×(m:k:d)×(i) (b: birey, g:grup, i:içerik, m:madde, k:konu ve d:düzye olmak üzere) deseni ile incelenmiştir. Analizde GENOVA ve her gruptan farklı katılımcı sayısı (b:g) nedeniyle dengelenmemiş desenlerde kullanılan urGENOVA programları kullanılmıştır. Çalışma sonucunda, gruptaki bireyler arasında farklılıkların olduğu, anketin gruplar arasındaki ayrımı gösterdiği bulunmuştur.

Jeon, Lee, Hwang ve Kang (2009), 90 okuldan 1477 lise 3. sınıf öğrencisine uygulanan Kore Dili, Matematik ve İngilizce başarı testlerinden oluşan Kore lise yetenek testinden elde edilen okul başarı puanlarının güvenilirliğini kestirmede çok düzeyli (multilevel) modeller ile genellenebilirlik kuramını kullanmışlardır. Çalışmanın amacı çok düzeyli modeller ve genellenebilirlik kuramı kullanarak dengeli ve dengelenmemiş verilerden elde edilen sonuçlara dayanarak okul puanlarının güvenilirliklerini karşılaştırmaktır. Çalışmada öğrencilerin okullarla yuvalandığı ve tüm öğrencilerin Kore Dili, Matematik ve İngilizce testlerindeki tüm maddeleri aldığı desen kullanılmıştır. Böylelikle toplam okul sayısı 90 olmak üzere, her bir okuldan gelen öğrenci sayılarının 10-20 arasında değiştiği dengelenmemiş ve her bir okuldan gelen öğrenci sayılarının 10 olarak sabitlendiği dengeli tasarımlar için ayrı ayrı analizler yapılmıştır. Genellenebilirlik kuramı ve çok düzeyli modellerle elde edilen dengelenmiş verilere ait sonuçlarda her iki yaklaşımla aynı varyans bileşenleri ve güvenilirlik kestirimleri elde edilmiştir. Fakat dengelenmemiş verilerle yapılan G kuramı ve çok düzeyli modellerle elde edilen kestirimler arasında farklılaşma olmuştur. Ayrıca G kuramıyla aynı dengelenmiş ve dengelenmemiş desenlere ait varyans kestirimleri ve aynı senaryolara ait güvenilirlik kestirimleri her iki desende de birbirine yakın çıkmıştır.

Zibrowski, Myers, Norman, Goldszmidt (2011) tarafından yapılan çalışmanın amacı klinik öğrenmeleri değerlendirmek için kullanılan araçların güvenilirliğinin test edilmesidir. Araştırmacılar çalışma kapsamında klinik değerlendirmeler için klinik öğretim etkinliğini değerlendirme aracı ve 15 maddeden oluşan poliklinik öğrenmeler için öğrenci değerlendirmesinden oluşan iki değerlendirme aracını kullanmışlardır. Dört veya daha fazla sorunun

cevaplanmadığı durumlar analiz dışı bırakılmıştır. Kullanılan her iki araç için ayrı ayrı toplam test puanlarının güvenilirliğini ve puanlayıcılar arası güvenilirliği belirlemek için dengelenmemiş desen ile genellenebilirlik kuramı analizi yapılmıştır. Analizlerde (d:ö) × m deseni (ö: öğretim elemanı, d: stajyer doktor ve m: madde olmak üzere) kullanılmıştır. Analiz sonuçlarına göre her iki test güvenilirliği 0,57 ve 0,60 bulunmuştur. Ayrıca puanlayıcılar için açıklanan varyansın yüksek olması nedeniyle puanlayıcılar arası güvenilirlik düşük bulunmuştur.

Yurt dışında genellenebilirlik kuramı ile ilgili yapılan çalışmalar incelendiğinde, gerçek durum verilerine uygun olan dengelenmemiş desenler ile genellenebilirlik kuramı analizlerinin yapıldığı, aynı çalışmada dengelenmiş ve dengelenmemiş desenlere ait analizlerin yer aldığı çalışmaların olduğu görülmektedir. Genellenebilirlik kuramı ile yapılan çalışmaların daha çok performans değerlendirme çalışmalarının güvenilirliğini belirlemede kullanılmasıyla birlikte test puanlarının güvenilirliğini belirlemede de kullanıldığı görülmektedir. Ayrıca genellenebilirlik kuramıyla dengelenmiş ve dengelenmemiş desenlere ait puanlama güvenilirlik kestirimlerinin karşılaştırıldığı çalışmalardan elde edilen bulgulara göre, güvenilirlik katsayılarının her iki desende de birbirlerine yakın çıkma eğiliminde olduğu görülmüştür.

BÖLÜM III

YÖNTEM

Bu bölümde, araştırmanın modeli, çalışma grubu, araştırma verileri ve verilerin analizine yer verilmiştir.

Araştırmanın Modeli

Bu araştırmada, puanlayıcıların dönüşümlü olarak öğrencilerin performanslarını değerlendirdiği tıp eğitimindeki 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 İntramuskuler Enjeksiyon istasyonuna giren ve değerlendirme formundaki bilgileri tam doldurulmuş birinci sınıftaki 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. 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 8 puanlayıcının puanlamalarından yararlanılmıştır. Öğrencilerin İntramuskuler Enjeksiyon yapma becerilerinin puanlanmasında görev alan bu puanlayıcılar Tıp Eğitimi ve Bilişimi Anabilim Dalı tarafından belirlenmiş olup puanlayıcıların hepsi Tıp alanından seçilmiştir.

Veriler ve Toplanması

Araştırmada, Hacettepe Üniversitesi Tıp Fakültesi, Tıp Eğitimi ve Bilişimi Anabilim Dalı'nın 2010-2011 akademik yılına ait 1. Dönem öğrencilerinin Yapılandırılmış Objektif Klinik Sınavındaki İntramuskuler Enjeksiyon yapma becerileri istasyonundan elde edilen veriler kullanılmıştır.

İntramuskuler Enjeksiyon istasyonunda öğrenciler sırayla maket üzerinde İntramuskuler 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 Tıp Eğitimi ve Bilişimi Anabilim Dalı tarafından bu istasyonda kullanılmak üzere hazırlanan ve 17 görevden oluşan İntramuskuler Enjeksiyon Yapma Becerisi değerlendirme formunda tanımlanan beceriler kapsamında değerlendirmektedir. Puanlayıcılar ilgili istasyona ait değerlendirme formuyla öğrencilerin ilgili beceriyi gösterip göstermemesine göre her bir görev için gözlendi (1), gözlenmedi (0) şeklinde puanlama yapmaktadır. Puanlayıcıların istasyonda kullandıkları İntramuskuler Enjeksiyon Yapma Becerisi değerlendirme formu Ek 1'de verilmiştir. 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. Böylece OSCE sınavı esas yapısı itibarıyla puanlayıcıların her birinin birbirinden farklı sayıda öğrenciyi puanlaması bakımından denglenmemiş veri yapısındadır.

Elde edilen araştırma verileriyle, Genellenabilirlik Kuramı analizleri için her bir puanlayıcının araştırmaya katılan öğrencilerden sadece bir kısmını 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. Araştırmada kullanılan (ö:p) x g (ö: öğrenci, p: puanlayıcı ve g: görev olmak üzere) deseni için oluşturulmuş dengelenmiş ve dengelenmemiş veri yapıları Çizelge 3 ve Çizelge 4'te gösterilmiştir.

Çizelge 3. (ö:p) x g Deseni Dengelenmiş Veri Yapısı Örneği

	Öğrenci	Görev					
		g1	g2	g3	g16	g17
Puanlayıcı 1 (n _ö =30)	ö1	x	x	x	x	x
	•	x	x	x	x	x
	ö30	x	x	x	x	x
Puanlayıcı 2 (n _ö =30)	ö31	x	x	x	x	x
	•	x	x	x	x	x
	ö60	x	x	x	x	x
Puanlayıcı 3 (n _ö =30)	ö61	x	x	x	x	x
	•	x	x	x	x	x
	ö90	x	x	x	x	x
Puanlayıcı 4 (n _ö =30)	ö91	x	x	x	x	x
	•	x	x	x	x	x
	ö120	x	x	x	x	x
Puanlayıcı 5 (n _ö =30)	ö121	x	x	x	x	x
	•	x	x	x	x	x
	ö150	x	x	x	x	x
Puanlayıcı 6 (n _ö =30)	ö151	x	x	x	x	x
	•	x	x	x	x	x
	ö180	x	x	x	x	x
Puanlayıcı 7 (n _ö =30)	ö181	x	x	x	x	x
	•	x	x	x	x	x
	ö210	x	x	x	x	x
Puanlayıcı 8 (n _ö =30)	ö211	x	x	x	x	x
	•	x	x	x	x	x
	ö240	x	x	x	x	x

Çizelge 3'te gösterildiği gibi dengelenmiş (ö:p) x g deseninde araştırma kapsamında kullanılan sekiz puanlayıcıdan her biri birbirleriyle eşit sayıda olmak üzere 30'ar öğrenciyi 17 görev doğrultusunda puanlamaktadır. Böylece çalışma kapsamında toplamda 240 öğrenci sekiz puanlayıcı tarafından puanlanmıştır.

Çizelge 4. (ö:p) x g Deseni Dengelenmemiş Veri Yapısı Örneği

	Öğrenci	Görev					
		g1	g2	g3	g16	g17
Puanlayıcı 1 ($n_{\text{ö}}=23$)	ö1	x	x	x	x	x
	•	x	x	x	x	x
	ö23	x	x	x	x	x
Puanlayıcı 2 ($n_{\text{ö}}=25$)	ö24	x	x	x	x	x
	•	x	x	x	x	x
	ö48	x	x	x	x	x
Puanlayıcı 3 ($n_{\text{ö}}=26$)	ö49	x	x	x	x	x
	•	x	x	x	x	x
	ö74	x	x	x	x	x
Puanlayıcı 4 ($n_{\text{ö}}=28$)	ö75	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	ö102	x	x	x	x	x
Puanlayıcı 5 ($n_{\text{ö}}=30$)	ö103	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	ö132	x	x	x	x	x
Puanlayıcı 6 ($n_{\text{ö}}=31$)	ö133	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	ö163	x	x	x	x	x
Puanlayıcı 7 ($n_{\text{ö}}=36$)	ö164	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	ö199	x	x	x	x	x
Puanlayıcı 8 ($n_{\text{ö}}=41$)	ö200	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	•	x	x	x	x	x
	ö240	x	x	x	x	x

Çizelge 4'te gösterildiği gibi dengelenmemiş (ö:p) x g deseninde 240 öğrenciden bu kez 23 öğrenci 1. puanlayıcı tarafından, 25 öğrenci 2. puanlayıcı tarafından, 26 öğrenci 3. puanlayıcı tarafından, 28 öğrenci 4. puanlayıcı tarafından, 30 öğrenci 5. puanlayıcı tarafından, 31 öğrenci 6. puanlayıcı tarafından, 36 öğrenci 7. puanlayıcı tarafından ve 41 öğrenci de 8. puanlayıcı tarafından 17 görev doğrultusunda puanlanmaktadır.

Çizelge 3 ve Çizelge 4'te görüldüğü gibi dengelenmiş ve dengelenmemiş desenlerde toplam 240 öğrenci, sekiz puanlayıcı tarafından aynı görevler doğrultusunda puanlanmıştır. Dengelenmiş ve dengelenmemiş desenlerin çalışma gruplarına ait betimsel istatistikler ise Çizelge 5'te verilmiştir.

Çizelge 5. Araştırma Kapsamında Kullanılan Dengelenmiş ve Dengelenmemiş (ö:p) x g Desenlerine Ait Betimsel İstatistikler

	<u>Dengelenmiş Desen</u>	<u>Dengelenmemiş Desen</u>
• Öğrenci Sayısı	240	240
• Görev Sayısı	17	17
• Puanlayıcı Sayısı	8	8
• Puanlayıcıların Her Birinin Puanladığı Öğrenci Sayısı	30, 30, 30, 30, 30, 30, 30, 30	23, 25, 26, 28, 30, 31, 36, 41
• Ortalama	16,02	15,99
• Standart Sapma	1,24	1,23
• Çarpıklık	-1,635	-1,613
• Basıklık	3,04	3,04

Çizelge 5'te görüldüğü gibi her iki desendeki toplam öğrenci sayıları 240 olmak üzere, bu öğrenciler iki durumda da aynı sekiz puanlayıcı tarafından ve aynı 17 İntramuskuler Enjeksiyon yapma becerisine ait görevler

doğrultusunda puanlanmıştır. Dengelenmiş desende öğrencilerin 17 görev doğrultusunda aldığı puanların ortalaması 16,02 iken, dengelenmemiş desende öğrencilerin aynı 17 görev kapsamındaki puan ortalaması 15,99'dur. Her iki desende öğrencilerin ilgili performans durumuna ait ortalamaları ve standart sapmaları birbirine oldukça yakındır. Bununla birlikte her iki desene ait çalışma grubunun çarpıklık ve basıklık değerleri birbirine çok yakın olmakla birlikte her iki gruba ait puanlar sola çarpık ve sivri bir dağılım göstermektedir. Bu durumda öğrencilerin performans durumuna ait başarılarının yüksek olduğu, ilgili istasyonun öğrencilere kolay geldiği, grubun ölçülen beceri bakımından homojen olduğu yorumu yapılabilir. Sonuç olarak, her iki desende öğrencilerin ilgili performans durumuna ait dağılımları benzeşiklik göstermektedir.

Araştırmada kullanılan veriler, Hacettepe Üniversitesi Tıp Fakültesi Tıp Eğitimi ve Bilişimi Anabilim Dalı Başkanlığının izni ile kullanılmıştır.

Verilerin Analizi

Verilerin analizinde, alt problemlere göre belirlenen dengelenmiş ve dengelenmemiş (ö:p) x g desenleri için ayrı ayrı genellenebilirlik kuramı analizleri yapılmış ve her iki durumdan elde edilen sonuçlar karşılaştırılmıştır.

Araştırmanın ilk kısmında sekiz puanlayıcıdan her birinin birbiriyle eşit sayıda öğrenciyi (30) dönüşümlü olarak 17 görev doğrultusunda puanlamasıyla oluşturulmuş dengelenmiş (ö:p) x g deseni için genellenebilirlik ve karar çalışması yapılmıştır. Araştırmanın ikinci kısmında ise ilk durumda kullanılan puanlayıcıların OSCE kapsamında puanladığı öğrenci sayıları da dikkate alınarak her bir puanlayıcının birbirlerinden farklı sayıda öğrenciyi, 17 görev doğrultusunda puanlamasıyla oluşturulmuş dengelenmemiş (ö:p) x g deseni için genellenebilirlik çalışması ve karar çalışması yapılmıştır. Her iki durumda da kullanılan görevler, puanlayıcılar ve puanlanan toplam öğrenci sayısı aynı kalmak şartıyla araştırmanın son kısmında, iki durum için yapılan G ve karar çalışmalarından elde edilen sonuçlar karşılaştırılmıştır.

Dengelenmiş desenlerle genellenebilirlik kuramı analizleri için GENOVA, EduG, SPSS gibi bilgisayar programları bulunmaktadır. Dengelenmemiş veri yapılarıyla genellenebilirlik kuramı analizleri geniş istatistiksel temeller gerektirmektedir. Fakat urGENOVA ve G_String gibi elektronik yazılımların ilerlemesi ile analizler kolaylaştırılmıştır.

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.; Bloch & Norman, 2011) ara yüz programı kullanılmıştır.

G_String programı Dr. Ralph Bloch ve Geoff Norman tarafından geliştirilmiştir. G_String ara yüz programı GENOVA syntax yazılımı sağlayan, dengelenmiş (balanced) ve dengelenmemiş (unbalanced) verilerle genellenebilirlik kuramı analizlerinde genellenebilirlik ve karar çalışmalarının yapılabileceği bir programdır.

Dengelenmiş ve dengelenmemiş (ö:p) x g desenleri için G_String ara yüz programıyla çalışma kapsamında yapılan G ve K çalışmalarına ait analiz çıktıları Ek 2 ve Ek 3'te verilmiştir.

BÖLÜM IV

BULGULAR VE YORUMLAR

Bu bölümde, araştırmanın genel amacı çerçevesinde elde edilen bulgular ve bulgulara ait yorumlara yer verilmiştir.

Dengelenmiş (ö:p) x g Deseninin Genellenebilirlik Kuramı Sonuçları

Öğrenci (ö) ve puanlayıcı (p) değişkenlerinin yuvalanmış, görev (g) değişkeninin ise çapraz tasarlandığı dengelenmiş (ö:p) x g deseninin genellenebilirlik ve karar çalışmalarından elde edilen bulgular aşağıda verilmiştir.

Dengelenmiş (ö:p) x g Deseniyle G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranları

OSCE sınavı, İntramuskuler Enjeksiyon yapma istasyonuna katılan birinci sınıf öğrencilerinden seçilen 240 öğrenci dönüşümlü olarak sekiz puanlayıcı tarafından enjeksiyon yapma becerilerine ait 17 görev doğrultusunda puanlanmıştır. Puanlayıcıların her birinin birbirleriyle eşit sayıda öğrenciyi (30 öğrenci) dönüşümlü olarak aynı görevler doğrultusunda puanlamasıyla oluşturulan dengelenmiş (ö:p) x g deseni (ö:öğrenci, p:puanlayıcı, g:görev) ile G çalışması yapılmıştır. Bu desenle yapılan G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları Çizelge 6'da verilmiştir.

Çizelge 6. Dengelenmiş (ö:p) x g Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranları

Varyans Kaynağı	Kareler Toplamı (S.S.)	Sd	Kareler Ortalaması (M.S.)	Varyans Bileşeni	%
g	8.62402	16	0.53900	0.00141	2.6
p	3.18824	7	0.45546	0.00043	0.8
ö : p	18.4549	232	0.07955	0.00199	3.6
gp	22.40343	112	0.20003	0.00514	9.4
gö : p,e	169.67843	3712	0.04571	0.04571	83.6
Toplam	222.34902	4079			100

ö: öğrenci, p: puanlayıcı, g: görev

Çizelge 6'da, dengelenmiş (ö:p) x g desenine ait G çalışması sonucu kestirilen varyans ve toplam varyansı açıklama oranları incelendiğinde, görev (g) ana etkisine ait varyansın 0.00141 olarak kestirildiği görülmektedir. Görev ana etkisi için kestirilen varyans, toplam varyansın %2.6'sını açıklamaktadır. Görevlere ait ana etki, bazı görevlerin diğerlerine göre daha zor olup olmadığını gösterir (Shavelson ve Webb, 1991). Görev ana etkisi için kestirilen varyans bileşeninin düşük olması, görevlerin zorluk-kolaylık düzeylerinin değişmediğini, bir sıra takibiyle yapılması gereken, ardışıklık gösteren görevlerin birbirlerine göre öğrencilere eşit güçlükte geldiğini göstermektedir.

Puanlayıcı etkisi için kestirilen varyans bileşeni 0.00043 çıkmıştır. Puanlayıcı ana etkisi için kestirilen varyans bileşeni, toplam varyansın %0.8'ini açıklamaktadır. Puanlayıcılar için kestirilen varyans bileşeni Çizelge 6'da görüldüğü gibi en düşük varyans değerine sahiptir. Böylece puanlayıcı değişkenliğinin öğrencilerin performansını etkilemediği ileri sürülebilir.

Puanlayıcıların öğrencileri dönüşümlü puanlaması nedeniyle öğrencilerin puanlayıcılarla yuvalandığı ö:p etkisine ait varyans bileşeni ($\sigma^2(\ddot{o}:p) = \sigma^2(\ddot{o},\ddot{o}p)$), öğrenci ana etkisine ($\sigma^2(\ddot{o})$) ve öğrenci x puanlayıcı ortak etkisine ($\sigma^2(\ddot{o}p)$) ait varyans bileşenlerinden oluşmaktadır (Brennan, 2001; Shavelson ve Webb, 1991). Çizelge 6'da gösterildiği gibi $\sigma^2(\ddot{o}:p)$ 0.00199 olarak kestirilmiştir. Bu varyans bileşeninin toplam varyansı açıklama oranı ise %3.6'dır. ö:p etkisine ait elde edilen varyans kestirimi ve toplam varyansı açıklama oranı yüksek değildir. Bu nedenle öğrencilerin ölçülen beceri bakımından farklılaşmadığı, öğrencilerin ortalama performanslarının birbirlerine göre farklılık göstermediği, öğrenci-puanlayıcı ortak etkileşiminin farklılaşmadığı, puanlayıcıların davranışlarının bir öğrenciden diğerine değişmediği ve puanlayıcıların kararlı puanlama yaptığı şeklinde yorumlanabilir.

Görev x puanlayıcı (gp) ortak etkisine ait varyans bileşeni 0.00514 olarak kestirilmiştir. Görev x puanlayıcı ortak etkisine ait varyans bileşeni, görev-puanlayıcı etkisinden kaynaklı farklılıkların olup olmadığını, puanlayıcıların bir görevden diğerine verdiği puanlardaki kararlılıklarını gösterir (Shavelson ve Webb, 1991). Çizelge 6'da görüldüğü gibi $\sigma^2(gp)$ varyans bileşeni toplam varyansın %9.4'ünü açıklama oranıyla diğer varyans bileşenleri içinde büyüklük bakımından ikinci sıradadır. Görev x puanlayıcı ortak etkisinin ve toplam varyansı açıklama oranının diğer varyans bileşenlerine göre daha yüksek olması, görev x puanlayıcı etkisinden gelen farklılıkların olduğu şeklinde yorumlanabilir.

Artık etkisine ait varyans bileşeni ($\sigma^2(g\ddot{o}:p,e)$) toplam varyansın %83.6'sını açıklayarak, toplam varyansı açıklama oranı bakımından en yüksek sırada yer almaktadır. Artık varyans bileşeni Eşitlik 17'de gösterildiği gibi öğrenci-görev ortak etkileşimi, üç yönlü etkileşim (öğrenci-görev-puanlayıcı) ve açıklanamayan diğer kaynakları içermektedir (Brennan, 2001; Shavelson ve Webb, 1991).

$$\sigma^2(g\ddot{o}:p,e) = \sigma^2(\ddot{o}g,\ddot{o}gp,e) = \sigma^2(\ddot{o}g) + \sigma^2(\ddot{o}gp,e) \quad (17)$$

Artık varyansa ait elde edilen bulgularda artık varyansın yüksek olması öğrenci-görev ortak etkileşiminin, üç yönlü etkileşimin ve/veya diğer açıklanamayan değişkenlik kaynaklarından (unexplained sources of variation) gelebilecek etkilerin yüksek olabileceğini göstermektedir.

Dengelenmiş (ö:p) x g Deseninde Puanlayıcı ve Öğrenci Sayılarının Artırılıp Azaltılması Senaryolarına Göre Yapılan Karar Çalışması

Puanlayıcıların her birinin birbirleriyle eşit sayıda öğrenciyi aynı görevler doğrultusunda dönüşümlü olarak puanladığı, bu nedenle de öğrencilerin (ö) puanlayıcılarla (p) yuvalandığı, görevlerin (g) ise çaprazlandığı dengelenmiş (ö:p) x g deseni ile yapılan karar çalışmasında, görevler ölçme konusu (object of measurement) olarak belirlenmiştir. Araştırmada kullanılan orijinal puanlayıcı sayısı ve her bir puanlayıcının puanladığı öğrenci sayısının artırılıp azaltılması senaryolarıyla yapılan karar çalışmasıyla kestirilen G ve Phi katsayılarına ait bulgular Çizelge 7’de verilmiştir.

Çizelge 7. Dengelenmiş (ö:p) x g Desenine Ait K Çalışması ile Puanlayıcı ve Öğrenci Sayılarının Değiştirilmesi Senaryolarına Göre Kestirilen G ve Phi Katsayıları

$n_{ö+}$	n_p	Dengelenmiş (ö: p) x g Deseni		
		$n_{ö:p}$	G	Phi
120	6	20, 20, 20, 20, 20, 20	0.533	0.515
120	8	15, 15, 15, 15, 15, 15, 15, 15	0.579	0.563
120	10	12, 12, 12, 12, 12, 12, 12, 12, 12, 12	0.612	0.596
240	6	40, 40, 40, 40, 40, 40	0.574	0.556
240	8	30, 30, 30, 30, 30, 30, 30, 30	0.629	0.612
240	10	24, 24, 24, 24, 24, 24, 24, 24, 24, 24	0.667	0.651
360	6	60, 60, 60, 60, 60, 60	0.589	0.571
360	8	45, 45, 45, 45, 45, 45, 45, 45	0.647	0.630
360	10	36, 36, 36, 36, 36, 36, 36, 36, 36, 36	0.687	0.672

$n_{ö+}$: Toplam öğrenci sayısı, n_p : Puanlayıcı sayısı, $n_{ö:p}$: Her bir puanlayıcının puanladığı öğrenci sayısı

Araştırma kapsamında kullanılan sekiz puanlayıcıdan her biri, dönüşümlü olarak 30 öğrenciyi 17 görev doğrultusunda puanlamıştır. Uygulamadaki puanlayıcı ($n_p:8$) ve her bir puanlayıcının puanladığı öğrenci sayısına ($n_{ö:p}:30$) göre yapılan karar çalışmasında G katsayısı 0.629, Phi katsayısı ise 0.612 olarak kestirilmiştir.

Çizelge 7'de puanlanan toplam öğrenci sayısı ($n_{ö+}: 240$) sabit kalmak koşuluyla puanlayıcı sayısının azaltılıp, altı puanlayıcının her birinin 40'ar öğrenciyi 17 görev doğrultusunda puanlamasıyla elde edilen G katsayısı 0.574, Phi katsayısı ise 0.556 olarak kestirilmiştir. Puanlayıcı sayısının azaltılması ($n_{ö+}:240$, $n_p:6$, $n_{ö:p}:40$) durumunda hesaplanan G ve Phi katsayıları, araştırmada kullanılan orijinal puanlayıcı ($n_p:8$) ve her bir puanlayıcının puanladığı öğrenci sayısı ile ($n_{ö:p}:30$) kestirilen G katsayısına göre 0.055, Phi katsayısına göre ise 0.056 azalmaktadır.

Puanlanan toplam öğrenci sayısı ($n_{ö+}: 240$) sabit kalmak koşuluyla bu kez puanlayıcı sayısının araştırmada kullanılan puanlayıcı sayısına göre artırılması, 10 puanlayıcının her birinin 24 öğrenciyi aynı görevler doğrultusunda puanlaması sonucu G katsayısı 0.667 ve Phi katsayısı da 0.651 olarak kestirilmiştir. Böylece puanlayıcı sayısının artırılması durumunda ($n_{ö+}: 240$, $n_p:10$, $n_{ö:p}:24$) elde edilen G katsayısı 8 puanlayıcıdan her birinin 30'ar öğrenciyi puanlamasıyla elde edilen G katsayısına göre 0.038, Phi katsayısı ise orijinal puanlayıcı ve öğrenci sayısına göre 0.039 artmaktadır.

Yukarıda elde edilen bulgular doğrultusunda, puanlanan toplam öğrenci sayısı ($n_{ö+}: 240$) sabit kalmak koşuluyla puanlayıcı sayısının artırılması ve buna bağlı olarak puanlayıcı başına düşen öğrenci sayısının azalması halinde G ve Phi katsayılarının yükseldiği görülmektedir. Aynı durum puanlanan toplam öğrenci sayısının 120 ve 360 olduğu senaryo durumlarında puanlayıcı sayısı değiştirildiğinde de geçerlidir. Fakat puanlanan toplam öğrenci sayısı sabitken puanlayıcı sayısının azaltılıp artırılması ile elde edilen artış miktarı azdır. Bu nedenle puanlayıcı sayısını artırmanın, G ve Phi katsayıları üzerinde büyük bir etki yaratmadığı yorumu yapılabilir. Bu durum Deliceoğlu (2009), Güler ve Gelbal (2010), Yılmaz Nalbantoğlu ve Gelbal

(2011) tarafından yapılan çalışmalarda da, puanlayıcılara ait varyans kestirimlerinin düşük çıkmasına bağlı olarak puanlayıcı sayısını artırmanın güvenilirliği çok artırmadığı bulgularıyla benzerlik göstermektedir.

Karar çalışması sonuçlarına göre araştırmada bu kez puanlayıcı sayısı ($n_{p:8}$) sabit kalmak koşuluyla, puanlanan toplam öğrenci sayısı azaltıldığında ($n_{ö+: 120}$, $n_{p:8}$, $n_{ö:p:15}$) G katsayısı 0.579, Phi katsayısı da 0.563 olarak kestirilmiştir. Aynı şekilde puanlayıcı sayısı sabitken, her bir puanlayıcının puanladığı öğrenci sayısı orijinal verilere göre artırıldığında ($n_{ö+: 360}$, $n_{p:8}$, $n_{ö:p:45}$) G katsayısı 0.647, Phi katsayısı da 0.630 olarak kestirilmiştir. Puanlayıcı sayısı sabit tutulup, öğrenci sayısının azaltılması durumunda araştırmada kullanılan verilere göre G katsayısı 0.05, Phi katsayısı 0.049 azalmaktadır. Puanlayıcı sayısı sabit tutulup, öğrenci sayısının artırılması durumunda ise G katsayısı araştırmada kullanılan verilere göre 0.018, Phi katsayısı da 0.018 artmaktadır. Bu sonuçlardan da görülebileceği gibi puanlayıcı sayısı sabitken öğrenci sayısının artırılması, buna bağlı olarak da puanlayıcı başına düşen öğrenci sayısının artması güvenilirliği fazla etkilememektedir. Ayrıca puanlayıcı başına düşen öğrenci sayısı arttıkça, bir başka deyişle her bir puanlayıcı daha çok öğrenciyi puanladıkça bu etki de giderek azalmaktadır.

Çizelge 7'de verilen senaryolardan bu kez araştırmada kullanılan verilere göre ($n_{ö+: 240}$, $n_{p:8}$, $n_{ö:p:30}$) puanlayıcı ve her bir puanlayıcının puanladığı öğrenci sayısının birlikte artırılması ($n_{ö+: 360}$, $n_{p:10}$, $n_{ö:p:36}$) durumunda G katsayısının 0.687, Phi katsayısının da 0.672 olarak hesaplanmıştır. Puanlayıcı ve öğrenci sayılarının birlikte artırılması durumunda G katsayısındaki artış uygulamadaki verilere göre 0.058, Phi katsayısındaki artış ise 0.060 olmaktadır.

Dengelenmiş ($ö:p$) x g deseni ile senaryolara göre yapılan karar çalışmasından elde edilen bulgular doğrultusunda, puanlayıcı ve/veya öğrenci sayısındaki artışa bağlı olarak G ve Phi katsayıları da giderek artmaktadır. Ancak, araştırmada kullanılan orijinal verilere göre puanlayıcı ve öğrenci sayılarının birlikte artırılmasının, tek tek artırılmasına oranla G ve Phi

katsayılarında daha yüksek bir artışa neden olduğu görülmektedir. Bu nedenle, OSCE sınavına katılan gerçek öğrenci ve puanlayıcı sayıları dikkate alındığında elde edilecek G ve Phi katsayılarının $n_{ö+}$: 360, n_p :10, $n_{ö:p}$:36 durumuna göre daha yüksek olması beklenmektedir. Ayrıca aynı senaryo durumu için kestirilen G katsayısının Phi katsayısından daha yüksek olduğu görülmektedir. Bunun nedeni, aynı senaryo durumu için G katsayısı hesaplanırken kullanılan bağıl hata varyansının Phi katsayısında kullanılan mutlak hata varyansından düşük olmasıdır. Bağıl ve mutlak hatalar için yapılan karar çalışmaları (Atılğan, 2004; Brennan, 2000, 2001; Güler, 2008; Nalbantoğlu, 2009; Yelboğa, 2007) göstermektedir ki; tüm senaryo durumlarına göre mutlak hata varyansları, bağıl hata varyanslarından daha yüksek kestirilmektedir. Ayrıca elde edilen bulgular doğrultusunda, senaryolara bağlı artış ve azalmada G ve Phi katsayılarının paralellik gösterdiği de görülmektedir.

Dengelenmemiş (ö:p) x g Desenin Genellenebilirlik Kuramı Sonuçları

Öğrenci (ö) ve puanlayıcı (p) değişkenlerinin yuvalanmış, görev (g) değişkeninin ise çapraz tasarlandığı dengelenmemiş (ö:p) x g deseninin genellenebilirlik ve karar çalışmalarından elde edilen bulgular aşağıda verilmiştir.

Dengelenmemiş (ö:p) x g Deseniyle G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranları

Dengelenmemiş (ö:p) x g deseniyle yapılan G çalışmasında, dengelenmiş (ö:p) x g deseninde kullanılan OSCE sınavı İntramuskuler Enjeksiyon yapma istasyonuna katılan 240 birinci sınıf öğrencisinin yine aynı puanlayıcılar tarafından aynı görevler doğrultusunda puanlanmasından elde edilen veriler kullanılmıştır. Bununla birlikte dengelenmemiş (ö:p) x g deseniyle yapılan G çalışmasının dengelenmiş (ö:p) x g deseninden farkı, her bir puanlayıcının bu kez birbirlerinden farklı sayıda öğrenciyi 17 görev

doğrultusunda puanlamasıyla G çalışması yapılmasıdır. Dengelenmiş (ö:p) x g deseninden farklı olarak dengelenmemiş desende 1. puanlayıcı 23, 2. puanlayıcı 25, 3. puanlayıcı 26, 4. puanlayıcı 28, 5. puanlayıcı 30, 6. puanlayıcı 31, 7. puanlayıcı 36 ve 8. puanlayıcı 41 öğrenciyi puanlamıştır. Bu desenle yapılan G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları ise Çizelge 8'de verilmiştir.

Çizelge 8. Dengelenmemiş (ö:p) x g Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranları

Varyans Kaynağı	Kareler Toplamı (S.S.)	Sd	Kareler Ortalaması (M.S.)	Varyans Bileşeni	%
g	8.90637	16	0.55665	0.00142	2.5
p	2.82273	7	0.40325	0.00032	0.6
ö : p	18.52800	232	0.07986	0.00193	3.4
gp	23.35529	112	0.20853	0.00541	9.6
gö : p,e	174.91481	3712	0.04712	0.04712	83.8
Toplam	228.52721	4079			100

ö: öğrenci, p: puanlayıcı, g: görev

Çizelge 8'deki dengelenmemiş (ö:p) x g desenine ait G çalışması sonucu kestirilen varyans ve toplam varyansı açıklama oranları incelendiğinde, görev (g) ana etkisine ait varyansın 0.00142 olarak kestirildiği görülmektedir. Görev ana etkisi için kestirilen varyans, toplam varyansın %2.5'ini açıklamaktadır. Görevlere ait ana etki, bazı görevlerin diğerlerine göre daha zor olup olmadığını gösterir (Shavelson ve Webb, 1991). Bu nedenle görev ana etkisi için kestirilen varyans bileşeninin düşük olması,

görevlerin zorluk- kolaylık düzeylerinin değişmediğini, görevlerin birbirlerine göre öğrencilere eşit güçlükte geldiğini göstermektedir.

Puanlayıcı ana etkisi için kestirilen varyans bileşeni toplam varyansın %0.6'sını açıklamaktadır. Puanlayıcılar için kestirilen varyans bileşeni Çizelge 8'de görüldüğü gibi en düşük varyans değerine sahiptir. Böylece puanlayıcı değişkenliğinin öğrencilerin performansını etkilemediği ileri sürülebilir.

Çalışmada, her bir puanlayıcının birbirlerinden farklı sayıda öğrenciyi dönüşümlü puanlaması nedeniyle öğrencilerin puanlayıcılarla yuvalandığı ö:p etkisine ait varyans bileşeni 0.00193 olarak kestirilmiştir. Bu varyans bileşeninin toplam varyans içindeki oranı ise %3.4'tür. Puanlayıcıların öğrencileri dönüşümlü puanlaması nedeniyle öğrencilerin puanlayıcılarla yuvalandığı ö:p etkisine ait varyans bileşeni ($\sigma^2(\ddot{o}:p) = \sigma^2(\ddot{o}, \ddot{o}p)$), öğrenci ana etkisine ($\sigma^2(\ddot{o})$) ve öğrenci x puanlayıcı ortak etkisine ($\sigma^2(\ddot{o}p)$) ait varyans bileşenlerinden oluşmaktadır (Brennan, 2001; Shavelson ve Webb, 1991). Bu nedenle ö:p etkisine ait varyans kestirimi ve toplam varyansı açıklama oranının düşük olması, öğrencilerin İntramuskuler Enjeksiyon yapma becerisi bakımından ortalama performanslarının birbirlerine göre farklılık göstermediği, grubun ölçülen özellik bakımından benzeşik olduğu, öğrenci-puanlayıcı ortak etkileşiminden kaynaklı farklılıkların olmadığı ve puanlayıcıların davranışlarının bir öğrenciden diğerine değişmediği şeklinde yorumlanabilir.

Görev x puanlayıcı (gp) ortak etkisine ait varyans bileşeni 0.00541 olarak kestirilmiştir. Görev x puanlayıcı ortak etkisine ait varyans bileşeni, görev-puanlayıcı etkisinden kaynaklı farklılıkların olup olmadığını, puanlayıcıların bir görevden diğerine verdiği puanlardaki kararlılıklarını gösterir (Shavelson ve Webb, 1991). $\sigma^2(gp)$ varyans bileşeninin toplam varyansın %9.6'sını açıklaması ve diğer varyans bileşenlerine göre yüksek kestirilmesi nedeniyle görev x puanlayıcı ortak etkisinden gelen farklılıkların olduğu şeklinde yorumlanabilir.

Artık etkisine ait varyans bileşeni ($\sigma^2(g\ddot{o}:p,e)$) toplam varyansın %83.8'ini açıklayarak, toplam varyansı açıklama oranı bakımından en yüksek

sırada yer almaktadır. Artık varyans bileşeninin yüksek olması öğrenci-görev ortak etkileşiminin, üç yönlü etkileşimin ve/veya diğer açıklanamayan değişkenlik kaynaklarından (unexplained sources of variation) gelebilecek etkilerin yüksek olabileceğini göstermektedir.

Dengelenmemiş (ö:p) x g Deseninde Puanlayıcı ve Öğrenci Sayılarının Artırılıp Azaltılması Senaryolarına Göre Yapılan Karar Çalışması

Puanlayıcıların her birinin birbirlerinden farklı sayıda öğrenciyi aynı görevler doğrultusunda dönüşümlü olarak puanladığı, bu nedenle de öğrencilerin (ö) puanlayıcılarla (p) yuvalandığı, görevlerin (g) ise çaprazlandığı dengelenmemiş (ö:p) x g deseni ile farklı senaryolara göre yapılan karar çalışmasında görevler ölçme konusu (object of measurement) olarak belirlenmiştir. Dengelenmemiş (ö:p) x g deseni ile puanlayıcı ve her bir puanlayıcının puanladığı öğrenci sayısının artırılıp azaltılmasıyla yapılan karar çalışmasıyla kestirilen G ve Phi katsayılarına ait bulgular Çizelge 9'da verilmiştir.

Araştırma kapsamında kullanılan sekiz puanlayıcıdan her biri dönüşümlü olarak bu kez birbirlerinden farklı sayıda öğrenciyi 17 görev doğrultusunda puanlamıştır. Orijinal puanlayıcı ($n_p:8$) ve araştırmada kullanılan her bir puanlayıcının puanladığı öğrenci sayısına göre dengelenmemiş (ö:p) x g deseni ile yapılan karar çalışmasında G katsayısı 0.613, Phi katsayısı ise 0.600 olarak kestirilmiştir.

Çizelge 9. Dengelenmemiş (ö:p) x g Desenine Ait K Çalışması ile Puanlayıcı ve Öğrenci Sayılarının Değiştirilmesi Senaryolarına Göre Kestirilen G ve Phi Katsayıları

<u>Dengelenmemiş (ö: p) x g Deseni</u>					
n_{ö+}	n_p	n_{ö:p}	ñ_p	G	Phi
120	6	16,18,20,21,22,23	5.92	0.520	0.507
120	8	10,12,14,15,15,17,18,19	7.73	0.562	0.549
120	8	13,13,15,15,15,15,17,17	7.93	0.568	0.556
120	10	8,9,10,11,12,12,12,14,15,17	9.55	0.592	0.580
240	6	33,38,40,41,43,45	5.95	0.562	0.548
240	8	23,25,26,28,30,31,36,41	7.73	0.613	0.600
240	8	26,28,30,30,30,30,33,33	7.96	0.618	0.605
240	10	15,17,20,22,24,26,27,28,30,31	9.56	0.648	0.636
360	6	55,58,60,60,62,65	5.98	0.578	0.565
360	8	38,40,43,45,47,48,49,50	7.94	0.636	0.623
360	8	41,43,45,45,45,45,48,48	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_{ö+} : toplam öğrenci sayısı, n_p: puanlayıcı sayısı, n_{ö:p}: her bir puanlayıcının puanladığı öğrenci sayısı, ñ_p: (n_{ö+}²) / Σn_{ö:p}²

Çizelge 9’da görüldüğü gibi puanlanan toplam öğrenci sayısı (n_{ö+}: 240) sabit kalmak koşuluyla puanlayıcı sayısının azaltılıp, altı puanlayıcının her birinin birbirlerinden farklı sayıda öğrenciyi 17 görev doğrultusunda puanlamasıyla elde edilen G katsayısı 0.562, Phi katsayısı ise 0.548 olarak kestirilmiştir. Puanlanan toplam öğrenci sayısı (n_{ö+}: 240) yine sabit kalmak koşuluyla bu kez puanlayıcı sayısının araştırmada kullanılan puanlayıcı sayısına göre artırılması (n_p:10) durumunda G katsayısı 0.648 ve Phi katsayısı da 0.636 olarak kestirilmiştir. Böylece puanlayıcı sayısının azaltılması durumunda hesaplanan G ve Phi katsayıları, uygulamadaki verilerle hesaplanan G katsayısına göre 0.051, Phi katsayısı ise 0.052 azalmaktadır. Puanlayıcı sayısının artırılması durumunda (n_{ö+}: 240, n_p:10) ise G katsayısı uygulamadaki verilere (n_{ö+}:240, n_p:8, ñ_p:7,73) göre 0.035, Phi

katsayısı ise uygulamadaki puanlayıcı ve öğrenci sayısına göre 0.036 artmaktadır. Bu doğrultuda, öğrenci sayısı ($n_{\delta+}$: 240) sabit kalmak koşuluyla puanlayıcı sayısını artırmanın G ve Phi katsayılarını fazlaca etkilemediği görülmektedir.

Yapılan karar çalışmasında bu kez puanlayıcı sayısı (n_p :8) sabit kalmak koşuluyla, puanlanan toplam öğrenci sayısının ve buna bağlı olarak da her bir puanlayıcının puanladığı öğrenci sayısının orijinal verilere göre azaltılıp artırılması durumlarındaki G ve Phi katsayıları incelenmiştir. Çizelge 9'da puanlayıcı sayısı sabit kalmak koşuluyla her bir puanlayıcının puanladığı öğrenci sayısı azaltıldığında ($n_{\delta+}$: 120, n_p :8, \check{n}_p :7,73) G katsayısı 0.562, Phi katsayısı da 0.549 olarak kestirilmiştir. Puanlayıcı sayısı sabitken her bir puanlayıcının puanladığı öğrenci sayısı orijinal verilere göre artırıldığında ise ($n_{\delta+}$: 360, n_p :8, \check{n}_p :7,94) G katsayısı 0.636, Phi katsayısı da 0.623 olarak kestirilmiştir. Böylece puanlayıcı sayısının sabit tutulup öğrenci sayısının azaltılması durumunda hesaplanan G ve Phi katsayıları orijinal verilere göre 0.051 azalmaktadır. Puanlayıcı sayısı sabit tutulup öğrenci sayısının artırılması durumunda ise G ve Phi katsayıları orijinal verilere göre 0.023 artmaktadır. Bu doğrultuda, puanlayıcı sayısı (n_p :8) sabitken puanlayıcıların her birinin dönüşümlü olarak puanladığı öğrenci sayısının artırılmasının G ve Phi katsayılarını fazlaca etkilemediği, elde edilen artışın az olduğu görülmektedir. Ayrıca elde edilen bu artış, puanlayıcı sayısı sabitken puanlayıcı başına düşen öğrenci sayısı arttıkça da giderek azalmaktadır.

Çizelge 9'da verilen senaryolara göre araştırmada kullanılan orijinal puanlayıcı ve puanlanan toplam öğrenci sayısının birlikte artırılması durumunda ($n_{\delta+}$: 360, n_p :10, \check{n}_p :9,36) G katsayısının 0.664, Phi katsayısının da 0.652 olarak kestirildiği görülmektedir. Bu bulgudan, araştırmada kullanılan verilere göre puanlayıcı ve öğrenci sayılarının birlikte artırılmasının G ve Phi katsayılarında sırasıyla 0.051 ve 0.052'lik artışa neden olduğu görülmektedir. Böylece OSCE sınavında kullanılan gerçek öğrenci ve puanlayıcı sayıları dikkate alındığında elde edilecek G ve Phi katsayılarının $n_{\delta+}$:360, n_p :10, \check{n}_p :9,36 durumuna göre daha yüksek olacağı yorumu yapılabilir.

Araştırmada sekiz puanlayıcının toplamda 240 öğrenciyi dönüşümlü olarak ve her birinin birbirinden farklı sayıda öğrenciyi belirlenen görevler doğrultusunda puanlamasıyla elde edilen G katsayısı Çizelge 9'da verildiği gibi 0.613, Phi katsayısı ise 0.600 olarak kestirilmiştir. Araştırmada kullanılan aynı sayıda puanlayıcının ($n_p:8$) yine toplamda 240 öğrenciyi olmak üzere bu kez her bir puanlayıcının puanladığı öğrenci sayıları arasındaki farklılığın ilk duruma göre azaltılması halinde ($n_{\delta+}: 240, n_p:8, \check{n}_p:7,96$) G katsayısı 0.618, Phi katsayısı ise 0.605 olarak hesaplanmıştır. Bu durum puanlayıcı ve toplam öğrenci sayısı aynı kalmak şartıyla puanlayıcıların puanladığı öğrenci sayısı bakımından puanlayıcılar arası değişkenliğin azaltıldığında G ve Phi katsayılarının yükseldiğini göstermektedir. Aynı durum Çizelge 9'da verilen 120 öğrencinin 8 puanlayıcı tarafından ve 360 öğrencinin 8 puanlayıcı tarafından puanlanması senaryolarında da görülmektedir.

Dengelenmemiş (ö:p) x g deseni ile yapılan karar çalışmasından elde edilen bulgular doğrultusunda genel olarak, senaryolara göre yapılan artış ve azalmada G ve Phi katsayılarının paralellik gösterdiği, aynı senaryo durumu için kestirilen G ve Phi katsayıları birlikte incelendiğinde, G katsayısının Phi katsayısından daha yüksek kestirildiği görülmektedir. Ayrıca puanlanan toplam öğrenci sayısı ve puanlayıcı sayısı sabit kalmak koşuluyla puanlayıcıların her birinin puanladığı öğrenci sayıları arasındaki farklılık arttıkça G ve Phi katsayılarının düştüğü tespit edilmiştir. Senaryo durumlarına göre, puanlayıcı ve/veya öğrenci sayısındaki artışa bağlı olarak G ve Phi katsayılarının giderek yükseldiği görülmektedir. Elde edilen G ve Phi katsayılarını artırmak için ise araştırmada kullanılan orijinal verilere göre puanlayıcı ve öğrenci sayılarının birlikte artırılmasının daha uygun olduğu ve puanlayıcıların her birinin puanladığı öğrenci sayıları arasındaki farklılığın (dengesizliğin) azaltılması gerektiği ifade edilebilir.

Dengelenmiş ve Dengelenmemiş (ö:p) x g Desenlerinden Elde Edilen Genellenebilirlik Kuramı Sonuçlarının Karşılaştırılması

Öğrenci (ö) ve puanlayıcı (p) değişkenlerinin yuvalanmış, görev (g) değişkeninin ise çapraz tasarlandığı dengelenmiş ve dengelenmemiş (ö:p) x g desenlerinin genellenebilirlik ve karar çalışmalarından elde edilen bulgulara ait karşılaştırmalar aşağıda verilmiştir.

Her İki Desenden Elde Edilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranlarının Karşılaştırılması

Dengelenmiş (ö:p) x g desenine ait varyans kestirimlerini belirlemede her bir puanlayıcı 17 görev doğrultusunda 30 öğrenciyi puanlayarak toplamda 240 öğrenci puanlanmıştır. Dengelenmemiş (ö:p) x g deseninde ise 1. puanlayıcı 23, 2. puanlayıcı 25, 3. puanlayıcı 26, 4. puanlayıcı 28, 5. puanlayıcı 30, 6. puanlayıcı 31, 7. puanlayıcı 36 ve 8. puanlayıcının 41 öğrenciyi olmak üzere toplamda 240 öğrencinin 17 görev doğrultusunda puanlamasıyla varyans kestirimleri elde edilmiştir. Aynı veriler kullanılarak puanlayıcıların her birinin birbirleriyle aynı ve birbirlerinden farklı sayıda öğrenciyi dönüşümlü olarak puanlamasına ait iki durumda (ö:p) x g desenine göre kestirilmiş varyans ve toplam varyansı açıklama oranları Çizelge 10'da gösterilmiştir.

Çizelge 10'daki verilere göre, görevlere ait varyans bileşenlerinin toplam varyansı açıklama oranı, dengelenmiş (ö:p) x g deseninde %2.6 ve dengelenmemiş (ö:p) x g deseninde %2.5 olarak hesaplanmıştır. Her iki desene görevlere ait kestirilen varyans bileşenlerinin düşük olduğu görülmektedir. Bu nedenle her iki desen için de görevlerin zorluk-kolaylık bakımından farklılaşmadığı, ardışıklık gösteren görevlerin birbirlerine göre öğrencilere eşit güçlükte geldiği yorumu yapılabilir.

Çizelge 10. Dengelenmiş ve Dengelenmemiş (ö:p) x g Desenlerinden Elde Edilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Oranları

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

Çizelge 10'da puanlayıcı ana etkisi için verilen varyans bileşeni dengelenmiş (ö:p) x g deseninde 0.00043, dengelenmemiş (ö:p) x g deseninde ise 0.00032'dir. Ayrıca her iki desenle kestirilen puanlayıcı etkisine ait varyans bileşeninin toplam varyansı açıklama oranı dengelenmiş desende %0.8, dengelenmemiş desende ise %0.6'dır. Her iki desen için kestirilen varyans değerlerinin ve toplam varyansı açıklama oranlarının düşük ve birbirine yakın olduğu görülmektedir. Bu bağlamda, her iki desen için puanlayıcı değişkenliğinin öğrencilerin performansını etkilemediği, puanlayıcıların puanlama bakımından farklılığa neden olmadığı yorumu yapılabilir.

Puanlayıcıların aynı görevler doğrultusunda öğrencileri dönüşümlü olarak puanlamasıyla tasarlanan dengelenmiş ve dengelenmemiş (ö:p) x g desenlerinde öğrencilerin puanlayıcılarla yuvalandığı (ö:p) değişkene ait varyans bileşeni, dengelenmiş desende $\sigma^2(\text{ö:p})= 0.00199$ (%3.6), dengelenmemiş desende ise $\sigma^2(\text{ö:p})= 0.00193$ (%3.4) olarak kestirilmiştir. Elde edilen sonuçlardan da görüldüğü gibi ö:p değişkenine ait varyans kestirimleri ve bu kestirimlerin toplam varyansı açıklama oranları her iki

desende de hemen hemen birbirine yakın ve düşük çıkmıştır. Bu durumda her iki desen için de öğrencilerin İntramusküler Enjeksiyon yapma becerileri bakımından farklılaşmadığı ve puanlayıcıların her birinin davranışlarının bir öğrenciden diğerine değişmediği yorumu yapılabilir. Öğrencilerin ilgili beceri bakımından farklılaşmadığına ait elde edilen bu bulgu Çizelge 5'te öğrencilerin ilgili istasyondaki becerilerine ait dağılımdan grubun ölçülen beceri bakımından homojen çıkmasıyla da birbirini destekler niteliktedir.

Dengelenmiş ve dengelenmemiş (ö:p) x g desenlerinde görev x puanlayıcı ortak etkileşimine ait varyans bileşenleri dengelenmiş desende $\sigma^2(gp)=0.00514$ (%9.4), dengelenmemiş desende ise $\sigma^2(gp)=0.00541$ (%9.6) olarak kestirilmiştir. Elde edilen bu bulgulardan her iki desende 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 için de görev x puanlayıcı ortak etkisinden kaynaklı farklılıkların olduğu yorumu yapılabilir.

Çizelge 10 incelendiğinde puanlayıcıların birbirleriyle eşit sayıda öğrenciyi puanladığı dengelenmiş (ö:p) x g deseninde artık varyansın ($\sigma^2(gö:p)$) toplam varyansı açıklama oranının %83.6 olduğu, puanlayıcıların birbirinden farklı sayıda öğrenciyi puanladığı dengelenmemiş (ö:p) x g deseninde ise artık varyansın ($\sigma^2(gö:p)$) 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 (unexplained sources of variation) 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 yorumu yapılabilir.

Genel olarak, dengelenmiş ve dengelenmemiş (ö:p) x g desenleri ile yapılan G çalışmaları sonunda kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları incelendiğinde, her iki desende kestirilen varyans

bileşenlerinin ve toplam varyansı açıklama oranlarının benzer sonuçlar ürettiği görülmektedir. Araştırmadan elde edilen bu bulgular, 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 desenlerle elde edilen varyans kestirimlerinin benzerlik gösterdiği sonuçlarıyla birbirini desteklemektedir. 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 çok benzer olduğu sonuçları elde edilmiştir. Ayrıca Jeon, Lee, Hwang ve Kang'ın (2009), çok düzeyli modeller ve genellenebilirlik kuramı kullanarak dengelenmiş ve dengelenmemiş verilerden elde edilen sonuçlara dayanarak okul puanlarının güvenilirliklerini karşılaştırdığı çalışmasında da G kuramıyla aynı dengelenmiş ve dengelenmemiş desenlere ait varyans kestirimlerinin de birbirine yakın çıktığı sonucu elde edilmiştir.

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

Dengelenmiş ve dengelenmemiş (ö:p) x g desenlerinde öğrenci ve puanlayıcı sayılarının artırılıp azaltılması senaryolarına göre yapılan karar çalışmalarından elde edilen G ve Phi katsayıları Çizelge 11'de karşılaştırılmıştır.

Çizelge 11'de görüldüğü gibi, 8 puanlayıcıdan her birinin 30'ar öğrenciyi ($n_{ö+}=240$) uygulamadaki görevler doğrultusunda puanlamasıyla dengelenmiş desenden elde edilen G katsayısı 0.629, Phi katsayısı 0.612 olarak kestirilmiştir. Dengelenmiş (ö:p) x g deseninde kullanılan aynı verilerle puanlayıcıların her birinin bu kez birbirlerinden farklı sayıda öğrenciyi puanlamasıyla dengelenmemiş desenden elde edilen G katsayısı 0.613, Phi katsayısı ise 0.600 olarak kestirilmiştir. Puanlayıcıların birbirleriyle eşit ya da

birbirinden farklı sayıda öğrenciyi puanlamasıyla oluşturulmuş desenlerden, dengelenmiş desendeki G katsayısı dengelenmemiş desene göre 0.016, Phi katsayısı ise dengelenmemiş desene göre 0.012 daha yüksek hesaplanmıştır. Böylece, uygulamadaki verilerle her iki desenden elde edilen katsayıların birbirine yakın sonuçlar ürettiği yorumu yapılabilir.

Çizelge 11. Dengelenmiş ve Dengelenmemiş (ö:p) x g Desenleri İle Senaryolara Göre Yapılan K Çalışmalarından Elde Edilen G ve Phi Katsayılarının Karşılaştırılması

<u>Dengelenmiş (ö:p) x g Deseni</u>				<u>Dengelenmemiş (ö:p) x g Deseni</u>				
$n_{ö+}$	n_p	G	Phi	$n_{ö+}$	n_p	\check{n}_p	G	Phi
				120	6	5.92	0.520	0.507
120	6	0.533	0.515	120	8	7.73	0.562	0.549
120	8	0.579	0.563	120	8	7.93	0.568	0.556
120	10	0.612	0.596	120	10	9.55	0.592	0.580
240	6	0.574	0.556	240	6	5.95	0.562	0.548
240	8	0.629	0.612	240	8	7.73	0.613	0.600
240	10	0.667	0.651	240	8	7.96	0.618	0.605
360	6	0.589	0.571	240	10	9.56	0.648	0.636
360	8	0.647	0.630	360	6	5.98	0.578	0.565
360	10	0.687	0.672	360	8	7.94	0.636	0.623
				360	8	7.98	0.637	0.624
				360	10	9.36	0.664	0.652

$n_{ö+}$: toplam öğrenci sayısı n_p : puanlayıcı sayısı, \check{n}_p : $(n_{ö+}^2) / \sum n_{ö:p}^2$

Çizelge 11’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 ($n_{ö+}$: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 ($n_{\delta+}:240$, $n_p:8$) daha düşük, dengelenmemiş desenden ($n_{\delta+}:240$, $n_p:8$, $\check{n}_p:7.73$) ise daha yüksek olduğu görülmektedir. Bu durum, puanlayıcıların puanladığı öğrenci sayısı bakımından, puanlayıcılar arası değişkenliğin azaltıldığında G ve Phi katsayılarının yükselme eğiliminde olduğunu göstermektedir.

Çizelge 11'de her iki desene göre verilen senaryo durumlarında bu kez uygulamadaki toplam öğrenci sayısı sabit tutularak ($n_{\delta+}:240$) puanlayıcı sayısı azaltıldığında ($n_p=6$) 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 ($n_p=10$) 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, 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 düşük olduğu tespit edilmiştir.

Çizelge 11'deki verilerden bu kez puanlayıcı sayısı sabit tutularak ($n_p=8$) öğrenci sayısı azaltıldığında ($n_{\delta+}:120$) dengelenmiş desende G katsayısı 0.579, Phi katsayısı ise 0.563 olarak kestirilmiştir. Aynı durumda dengelenmemiş desende ($n_{\delta+}: 120$, $n_p:8$, $\check{n}_p:7.73$) G katsayısı 0.562, Phi katsayısı da 0.549 olarak kestirilken; puanlayıcıların puanladığı öğrenci sayısı bakımından dengesizlik azaltıldığında ($n_{\delta+}: 120$, $n_p:8$, $\check{n}_p:7.93$) ise G katsayısı 0.568, Phi katsayısı da 0.556 olarak kestirilmiştir. Puanlayıcı sayısı uygulamadaki gibi sabit tutulup öğrenci sayısı artırıldığında ($n_{\delta+}:360$) ise dengelenmiş desene göre G katsayısı 0.647, Phi katsayısı 0.630 olarak kestirilken; dengelenmemiş desene göre ($n_{\delta+}:360$, $n_p:8$, $\check{n}_p:7.94$) ise G katsayısı 0.636, Phi katsayısı da 0.623 olarak kestirilmiştir. Ayrıca puanlayıcıların puanladığı öğrenci sayısı bakımından dengesizlik azaltıldığında ($n_{\delta+}: 360$, $n_p:8$, $\check{n}_p:7.98$), G katsayısı 0.637, Phi katsayısı da 0.624 olarak kestirilmiştir. Böylece, puanlayıcı sayısı sabit tutularak puanlanan

toplam öğrenci sayısının azaltılması ve artırılması durumlarında G ve Phi katsayılarının dengelenmiş (ö:p) x g deseninde yüksek çıkma eğiliminde olmakla birlikte her iki desene elde edilen katsayıların birbirlerine yakın olduğu yorumu yapılabilir. Ayrıca farklı senaryo durumlarına göre dengelenmemiş durumdan dengelenmiş duruma gidildikçe G ve Phi katsayılarının yükselme eğiliminde olduğu görülmektedir.

Her iki desende puanlayıcı ve öğrenci sayısının uygulamadaki verilere göre birlikte artırılması durumunda ($n_{ö+}$: 360, n_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 her iki desende de arttığı, her iki desene elde edilen katsayıların ise birbirine yakın sonuçlar ürettiği yorumu yapılabilir.

Çizelge 11’de gösterildiği gibi 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 katsayılardan daha yüksek çıkmıştır. Aynı durum dengelenmemiş desende 360 öğrencinin 8 puanlayıcı tarafından puanlanması ile elde edilen katsayıların dengelenmiş desende 120 öğrencinin 8 puanlayıcı tarafından puanlanmasıyla elde edilen katsayılarda da görülmektedir. Bu durum dengelenmemiş veri yapılarından veri indirgemesi yapılarak dengeli desenlerle güvenilirlik katsayıları hesaplandığında çok fazla veri kaybı olduğunda güvenilirliğin düştüğünü göstermektedir.

Genel olarak elde edilen bulgulardan, her iki desende tüm senaryo durumlarına göre G katsayılarının Phi katsayılarından daha yüksek kestirildiği, puanlayıcı ve öğrenci sayılarının artırılmasının her iki desende de G ve Phi katsayılarını artırdığı, dengelenmemiş yapıdan dengeli yapıya doğru gidildikçe güvenilirliğin artma eğiliminde olduğu görülmektedir. Bununla birlikte, toplam

veri sayısı aynı kalmak şartıyla senaryolara göre dengelenmiş ve dengelenmemiş (ö:p) x g desenleri ile kestirilen katsayılar arasındaki farklılığın az olduğu tespit edilmiştir. Böylece her iki desende de senaryolara göre yakın sonuçlar elde edildiği yorumu yapılabilir. Ayrıca araştırmadan elde edilen bu bulgular, Jeon, Lee, Hwang ve Kang (2009), Lee ve Frisbie (1999), Malhotra ve Sharma (2008), Shavelson ve Webb (1981), tarafından dengelenmiş ve dengelenmemiş desenlerle farklı senaryo durumlarına göre farklı çalışmalarla elde edilen güvenilirlik katsayılarının her iki desende de birbirlerine yakın sonuçlar verdiği bulgularıyla da tutarlılık göstermektedir.

BÖLÜM V

SONUÇ VE ÖNERİLER

Bu bölümde, araştırmadan elde edilen sonuçlara ve bu sonuçlara dayalı olarak sunulan önerilere yer verilmiştir.

Sonuçlar

Bu araştırmada, psikomotor becerilerin ağırlıkta olduğu performans değerlendirme çalışmasında puanlayıcıların her birinin birbiriyle eşit (dengelenmiş) ve birbirinden farklı sayıda (dengelenmemiş) öğrenciyi dönüşümlü olarak puanlanmasına ait iki durumda genellenebilirlik kuramına göre elde edilen G ve K çalışmaları sonuçlarının karşılaştırılması amaçlanmıştır. Bu karşılaştırma yapılırken, puanlayıcıların birbirlerine eşit ya da birbirlerinden farklı sayıda öğrenciyi puanlamasının güvenilirlik katsayılarına etkisi de incelenmiştir.

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

1. Her iki desede de ardışıklık gösteren görevlerin öğrencilere eşit güçlükte geldiği sonucuna varılmıştır.

2. Her iki desen için puanlayıcı değişkenliğinin öğrencilerin performansını etkilemediği, puanlayıcıların puanlama bakımından farklılığa neden olmadığı sonucu elde edilmiştir.

3. Her iki desende de öğrencilerin İntramusküler Enjeksiyon yapma becerileri bakımından farklılaşmadığı, bireylerin ilgili beceri bakımından homojen olduğu ve puanlayıcıların her birinin davranışlarının bir öğrenciden diğerine değişmediği tespit edilmiştir.

Sonuç olarak, İntramusküler Enjeksiyon yapma istasyonu için puanlayıcıların her birinin birbiriyle eşit ya da birbirinden farklı sayıda öğrenciyi puanlamasıyla oluşturulmuş dengelenmiş ve dengelenmemiş (ö:p)xg desenleri ile yapılan G çalışmaları sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama oranları her iki desende de birbirine yakın sonuçlar üretmektedir.

OSCE sınavı İntramusküler Enjeksiyon yapma istasyonu için K çalışmasıyla senaryolara göre dengelenmiş ve dengelenmemiş (ö:p) x g desenleri ile elde edilen G ve Phi katsayılarına ait sonuçlar ise aşağıda verilmiştir.

1. Araştırmada kullanılan 240 öğrencinin sekiz puanlayıcı tarafından aynı görevler doğrultusunda puanlanmasında, puanlayıcıların puanladıkları öğrenci sayısının birbirlerine eşit ve birbirlerinden farklı olması durumlarına göre oluşturulmuş iki desenden, puanlayıcıların her birinin eşit sayıda öğrenciyi puanladığı dengelenmiş desende G katsayısı 0.629, Phi katsayısı ise 0.612 olarak kestirilmiştir. Puanlayıcıların her birinin birbirlerinden farklı sayıda öğrenci puanladığı dengelenmemiş desende ise G katsayısı 0.613, Phi katsayısı 0.600 olarak kestirilmiştir. Bu doğrultuda, uygulamadaki verilerle her iki desene göre hesaplanan katsayıların birbirine yakın sonuçlar ürettiği sonucuna varılmıştır.

2. Öğrenci ve puanlayıcı sayılarının artırılıp azaltılmasıyla yapılan karar çalışmalarında;

a. Toplam veri sayısı her iki desende de aynı kalmak şartıyla farklı senaryo durumlarına göre elde edilen G ve Phi katsayılarının, her iki desende

birbirine yakın çıkmakla birlikte katsayıların dengelenmemiş durumdan dengelenmiş duruma gidildikçe yükselme eğiliminde olduğu tespit edilmiştir.

b. Dengelenmemiş veri yapılarından veri kaybı ile dengelenmiş veri yapısı üretip güvenilirlik hesaplandığında güvenilirliğin düştüğü sonucuna varılmıştır.

3. Performans değerlendirmenin İntramusküler Enjeksiyonu istasyonuna ait sonuçlarında senaryolara göre kestirilen Phi katsayılarının G katsayılarından düşük olduğu görülmektedir. Fakat bu durumun G ve Phi katsayılarının kestirilmesinde kullanılan bağıl ve mutlak hata varyansları arasındaki farklılıktan, mutlak hata varyanslarının, bağıl hata varyanslarından daha yüksek kestirilmesinden kaynaklandığı bilinmektedir (Brennan, 2001).

4. Yapılan karar çalışmasında her iki desende elde edilen G ve Phi katsayılarının yüksek olmadığı görülmüştür. Farklı senaryo durumlarına göre yapılan karar çalışmalarında ise elde edilen bu katsayıların puanlayıcı ve puanlayıcıların her birini puanladığı öğrenci sayısına bağlı olarak yükseldiği tespit edilmiştir. Bununla birlikte her iki desende de puanlayıcılar ve öğrenciler için kestirilen varyansların toplam varyansları açıklama oranının düşük olması nedeniyle puanlayıcı veya öğrenci sayılarının tek başına artırılması yerine birlikte artırılmasının daha fazla artış sağladığı sonucuna varılmıştır.

5. Genellenebilirlik kuramında genellenebilirlik katsayısı ve Phi katsayısı, evren puanı varyansının gözlenen puan varyansına oranıdır. Gözlenen puan varyansı da evren puanı varyansı ve hata varyansının toplamına eşittir (Brennan, 2001; Shavelson ve Webb, 1991). Grup benzeşikliği arttıkça varyans kestirimleri azalmaktadır (Baykul, 2000). Evren puanı varyansının azalmasına bağlı olarak da elde edilecek güvenilirlik katsayısı düşmektedir. Araştırmadan elde edilen sonuçlardan, desenlere göre hesaplanan G ve Phi katsayılarının görevlere ait varyans bileşeninin düşük çıkması, görevlerin zorluk ve kolaylık bakımından farklılaşmaması, görevlerin güçlüklerinin birbirine yakın olması ve öğrencilerin görevlerdeki performanslarının farklılık göstermemesi, grubun ölçülen özellik bakımından homojen olması nedenlerine bağlı olarak daha düşük çıktığı yorumu yapılabilir.

Yukarıda yapılan tüm açıklamalar ışığında, puanlayıcıların dönüşümlü olarak birbirleriyle aynı ya da birbirlerinden farklı sayıda öğrenciyi aynı görevler doğrultusunda puanlamasından elde edilen varyans kestirimleri benzerlik göstermektedir. Toplam veri sayısı aynı kalmak şartıyla her iki desende de farklı puanlayıcı ve öğrenci senaryolarına göre kestirilen katsayılar arasında çok fark olmadığı, dengelenmiş ve dengelenmemiş desenlerle farklı senaryo durumlarına göre elde edilen güvenilirlik katsayılarının her iki desende de birbirlerine yakın olduğu belirlenmiştir.

Öneriler

Bu bölümde, araştırma bulgularına dayalı olarak geliştirilen ve ileride yapılabilecek başka araştırmalar için önerilere yer verilmiştir.

1. 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. Fakat her iki desen arasında veri kaybının olduğu, dengelenmemiş veri yapılarından veri indirgemesi yapılarak dengeli desenler elde edildiğinde kestirilen güvenilirlik katsayıları arasındaki farklılığın arttığı bilinmektedir. Bu nedenle gerçek koşullarda dengelenmemiş veri yapısında olan durumlarda güvenilirlik hesaplanırken, dengelenmemiş desenlerden fazla veri kaybıyla dengeli yapılar oluşturarak güvenilirlik hesaplamak yerine, analizlerde kullanılan yazılımların geliştirilmesiyle dengelenmemiş desenlerle analiz yapmak daha uygun olabilir.

2. Araştırma kapsamında kullanılan OSCE sınavının genel yapısına uygunluğu nedeniyle genellenebilirlik kuramıyla güvenilirlik analizlerinde dengelenmemiş desenler kullanılması önerilmektedir.

3. Görevlerin güçlük bakımından farklı olmaması ve grubun ilgili beceriye ait görevlerdeki performanslarının benzeşik olması, elde edilen güvenilirlik katsayılarını etkilemiştir. Bu nedenle güvenilirliği artırmak için ilgili beceriye ait daha karmaşık görevler eklenebilir.

4. Bireylerin görevlere ait performansları benzeşik olduğundan elde edilen güvenilirlik katsayıları yorumlanırken dikkat edilmelidir.

5. Elde edilen G ve Phi katsayılarını artırmak için sınavda kullanılan puanlayıcı sayısı ve her bir puanlayıcıya düşen öğrenci sayısı artırılabilir. Fakat puanlayıcılara düşen öğrenci sayısının artması güvenilirliği artırmakla birlikte, her bir puanlayıcı daha çok öğrenciyi puanladıkça bu artış oranı giderek azalmaktadır. Bu nedenle puanlayıcıların iş gücü ve bununla bağlantılı olarak sıkılma durumları da dikkate alınarak, her bir puanlayıcıya çok sayıda öğrenci düşmeyecek şekilde öğrenci ve puanlayıcı dengesinin belirlenmesine dikkat edilmelidir.

6. 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 İnamusküler Enjeksiyon yapma ve diğer istasyonlarda öğrencilerin performanslarının puanlanmasında puanlayıcıların puanladığı öğrenci sayısı bakımından değişkenliğin fazla olmamasına dikkat edilmelidir.

7. Yapılacak çalışmalarda dengelenmemiş veri yapısı içeren her durumda değişkenin her bir koşulu arasındaki farklılık azaldığında güvenilirliğin arttığı ilkesine dikkat edilerek veri toplanabilir.

8. Bir başka çalışmada daha karmaşık becerilerin olduğu farklı performans durumlarında, farklı değişkenlik kaynakları kullanılarak dengelenmiş ve dengelenmemiş desenlerin karşılaştırılması yapılabilir.

9. Çalışmada OSCE sınavına ait İntramusküler Enjeksiyon yapma istasyonu ele alınmıştır. Başka çalışmalarda OSCE sınavına ait farklı klinik becerilerin değerlendirildiği istasyonlar ile aynı karşılaştırma yapılabilir.

10. Araştırmada yapılan karşılaştırma toplam veri sayısının eşit olduğu durumda yapılmıştır. Aynı karşılaştırma her iki desendeki toplam veri sayısı arasında farklılık olduğu durumlar için de tekrarlanabilir.

11. Araştırmadaki İntramusküler Enjeksiyon yapma becerileri ölçülen grup ilgili beceri bakımından homojen çıkmıştır. Bu nedenle bir başka çalışmada heterojen bir grup üzerinde dengelenmiş ve dengelenmemiş desenlerden elde edilen sonuçlar karşılaştırılabilir.

12. Bir başka çalışmada dengelenmemiş durumlardan dengelenmiş veri yapısı elde etmek için büyük veri indirgemesi yapılan çeşitli durumlardan çıkan sonuçların karşılaştırılması yapılabilir.

KAYNAKÇA

Akyüz, C., Elçin, M., Turan, S., Odabaşı, O., Abay, E., Onan, A., ve Başusta, B. (2011). *Hacettepe Üniversitesi 2011-2012 Öğretim Yılı İyi Hekimlik Uygulamaları*. Ankara: Hacettepe Üniversitesi Hastaneleri Basımevi.

Airasian, P.W.(1994). *Classroom Assessment*. New York: McGraw-Hill.

Arias, R. M. (2010). Performance Assessment. *Papeles del Psicologo*, 31(1), 85-96.

Arter, J. A., and Stiggins, R. J. (1992, April). *Performance Assessment in Education*. Paper Presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. Web: <http://www.eric.ed.gov/PDFS/ED346157.pdf> adresinden 25 Ağustos 2011'de alınmıştır.

Atılğan, H. (2004). *Genellenebilirlik Kuramı Ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılmasına İlişkin Bir Araştırma*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.

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.

Barnes, L. L. B., and Barnes, M. W. (1993). Academic Discipline and Generalizability of Student Evaluations of Instruction. *Research in Higher Education*, 34, 135-49.

- Barneveld, C. (2005). The Dependability of Medical Students' Performance Ratings as Documented on In Training Evaluations. *Academic Medicine*, 80(3), 309-312.
- Baykul, Y. (2000). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: ÖSYM Yayınları.
- Bekiroğlu, F. (2008). Performansa Dayalı Ölçümler: Teori ve Uygulama. *Türk Fen Eğitim Dergisi*, 5(1), 113- 131.
- Bell, J. F. (1985). Generalizability Theory: The Software Problem. *Journal of Educational Statistics*, 10(1), 19-29.
- Bergus, G. R., and Kreiter, C. D. (2007). The Reliability of Summative Judgements Based on Objective Structured Clinical Examination Cases Distributed Across the Clinical Year. *Medical Education*, 41, 661–666.
- Bloch, R., and Norman, G. (2011). G String 4 User Manual. Web: http://fhsperd.mcmaster.ca/g_string/download/g_string_4_manual.pdf adresinden 10 Ağustos 2011'de alınmıştır.
- Brennan, R. L., and Kane, M. T. (1977). An Index of Dependability for Mastery Tests. *Journal of Educational Measurement*, 14, 227-289.
- Brennan, R. L. (1992). *Elements of Generalizability Theory*. Iowa City: American College Testing.
- Brennan, R. L. (2000). Performance Assessments From The Perspective of Generalizability Theory. *Applied Psychological Measurement*, 24(4), 339-353.
- Brennan, R. L. (2001b). urGENOVA 2.1. Iowa City, IA: The University of Iowa. Web:<http://www.education.uiowa.edu/centers/casma/computer-programs.aspx#genova> adresinden 12 Temmuz 2011'de alınmıştır.

- Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlog.
- Brennan, R. L., Yin, P. and Kane, M. T. (2003). Methodology for Examining The Reliability of Group Mean Difference Scores. *Journal of Educational Measurement*, 40(3), 207-230.
- Briesch, A. M., Chafaouleas, S. M., and Riley-Tillman, T. C. (2010). Generalizability and Dependability of Behavior Assessment Methods to Estimate Academic Engagement: A Comparison of Systematic Direct Observation and Direct Behavior Rating. *School Psychology Review*, 39(3), 408-421.
- Brualdi, A. (1998). Implementing Performance Assessment in The Classroom. *Practical Assessment, Research & Evaluation*, 6(2). Web: <http://PAREonline.net/getvn.asp?v=6&n=2> adresinden 29 Ocak 2009'da alınmıştır.
- Cardinet, J., Tourneur, Y., and Allal, L. (1976). The Symmetry of Generalizability Theory: Applications to Educational Measurement. *Journal of Educational Measurement*, 13(2), 119-135.
- Cardinet, J., Tourneur, Y., and Allal, L. (1981). Extension of Generalizability Theory and Its Applications in Educational Measurement. *Journal of Educational Measurement*, 18(4), 183-204.
- Chang, L., and Hocevar, D. (2000). Models of Generalizability Theory in Analyzing Existing Faculty Evaluation Data. *Applied Measurement In Education*, 13(3), 255-275.
- Clauser, B. E., Swanson, D. B., and Clyman, S. G. (1996). The Generalizability of Scores from a Performance Assessment of Physicians' Patient Management Skills. *Academic Medicine*, 71(10), 109-111.

- Clauser, B. E., Harik, P., Margolis, M. J., Mee, J., and Swygert, K. (2008). The Generalizability of Documentation Scores from the USMLE Step 2 Clinical Skills Examination. *Academic Medicine*, 83(10), 41-44.
- Crocker, L., Algina, J. (1986). *Introduction to Classical And Modern Test Theory*. New York: Holt, Rinehart and Winston.
- Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N. (1972). *The Dependability of Behavioral Measurements: Theory of Generalizability for Scores and Profiles*. New York: Wiley.
- Cronbach, L. J., Linn, R. L., Brennan, R. L., and Haertel, E. H. (1997). Generalizability Analysis for Performance Assessments of Student Achievement or School Effectiveness. *Educational and Psychological Measurement*, 57(3), 373-399.
- Deliceoğlu, G. (2009). *Futbol Yetilerine İlişkin Dereceleme Ölçeğinin Genellenabilirlik ve Klasik Test Kuramına Dayalı Güvenirliklerinin Karşılaştırılması*. Yayınlanmamış doktora tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Dolmans, D.H.J.M., Wolfhagen, I.H.A.P, Schmidt, H.G., and Van Der Vleuten, C.P.M. (1994). A Rating Scale for Tutor Evaluation in a Problem Based Curriculum: Validity and Reliability. *Medical Education*, 28, 550-558.
- Dory, V., Gagnon, R., and Charlin, B. (2010). Is Case-Specificity Content-Specificity? An Analysis of Data From Extended-Matching Questions. *Advances in Health Sciences Education*, 15, 55-63.
- Durvasula, S., Netemeyer, R.G., Andrews, J. C., and Lysonski, S. (2006). Examining The Cross-Nationala Applicability of Multi-Item, Multi-Dimensional Measures Using Generalizability Theory. *Journal of International Business Studies*, 37, 469-483.

- Eason, S. H. (1989). Why Generalizability Theory Yields Better Results than Classical Test Theory. *Mid- South Educational Research Association Annual Meeting: 8-10 November 1989- Little Rock, AR*. Web: <http://eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED314434> adresinden 12 Eylül 2011'de alınmıştır.
- Elçin, M., Odabaşı, O., ve Sayek, İ. (2005). Yapılandırılmış Objektif Klinik Sınavlar. *Hacettepe Tıp Dergisi*, 36, 1-2.
- Fitzpatrick, A. R., Lee, G., and Gao, F. (2001). Assessing the Comparability of School Scores Across Test Forms That Are Not Parallel. *Applied Measurement in Education*, 14: (3), 285 -306.
- Foy, P. (2005). Estimating and Interpreting Variance Components in International Comparative Studies in Education. *Studies in Educational*, 31, 173-191.
- Gillmore, G. M., Kane, M. T., and Naccarato, R. W. (1978). The Generalizability of Student Ratings: Estimation of The Teacher and Course Components. *Journal of Educational Measurement*, 15,1, 1-13.
- Güler, N. (2008). *Klasik Test Kuramı, Genellenebilirlik Kuramı ve Rasch Modeli Üzerine Bir Araştırma*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Güler, N. ve Çetin, B. (2010). *Genellenebilirlik Kuramı ve Klasik Test Kuramı Açısından Gözetmen Sayısı Etkisinin İncelenmesi*. II. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sunuldu, Mersin.
- 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.
- Harden, R. M., and Gleeson, F. A. (1979). Assessment of Clinical Competence Using An Objective Structured Clinical Examination (OSCE). *Medical Education*, 13 (1):41–54.
- Jarjoura, D., and Brennan, R. L. (1981). *Three Variance Components Models for Some Measurement Procedures in Which Unequal Numbers of Items Fall into Discrete Categories* (ACT Tech. Bulletin No. 37).Iowa City, IA: American College Testing.
- Jarjoura, D., and Early, L., Androulakakis, V. (2004). A Multivariate Generalizability Model For Clinical Skills Assessments. *Educational and Psychological Measurement*, 64(1), 22-39.
- Jeon, M. J., Lee, G., Hwang, J. W, and Kang, S. J. (2009). Estimating Reliability of School-Level Scores Using Multilevel and Generalizability theory Models. *Asia Pacific Education Rev.*, 10, 149-158.
- Kane, M. T., and Brennan, R. L. (1977). The Generalizability of Class Means. *Review of Educational Research*, 47, 267–292.
- Kaufman, J., and Schering, A. (2007). Analysis of Variance ANOVA. Wiley Encyclopedia of Clinical Trials, John Wiley & Sons, Inc.
- Khatti, N., Reeve, A., and Kane, M. (1998). *Principles and Practices of Performance Assessment*. Mahwah, NJ: Lawrence Erlbaum.
- Kraiger, K. (1989). Generalizability Theory: An Assessment of Its Relevance To The Air Force Job Performance Measurement Project. Interim Technical Paper. Web: <http://www.dtic.mil/dtic/tr/fulltext/u2/a207107.pdf> adresinden 25 Eylül 2011'de alınmıştır.

- Kreiter, C. D., and Ferguson, K. J. (2001). Examining the Generalizability of Ratings across Clerkships Using a Clinical Evaluation Form. *Evaluation The Health Professions*, 24 (1), 36-46.
- Kutlu, Ö., Doğan, C.D., ve Karakaya, İ. (2010). *Öğrenci Başarısının Belirlenmesi Performansa ve Portfolyoya Dayalı Durum Belirleme*. Ankara: Pegem Akademi.
- Lane, S., and Sabers, D. (1989). Use of Generalizability Theory for Estimating the Dependability of a Scoring System for Sample Essays. *Applied Measurement In Education*, 2(3),195-205.
- Lee, G., and 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.
- Lee, G., and Fitzpatrick (2003). The Effect of a Student Sampling Plan on Estimates of the Standart Errors for Student Passing Rates. *Journal of Educational Measurement*, 40(1), 17-28.
- Lei, P., Smith, M., and Suen, H. K. (2007). The Use of Generalizability Theory To Estimate Data Reliability in Single Subject Observational Research. *Psychology in the Schools*, 44(5), 433-439.
- Malhotra, M. K., and Sharma, S. (2008). Measurement Equivalence Using Generalizability Theory: An Examination of Manufacturing Flexibility Dimensions. *Decision Sciences*, 39(4), 643-669.
- Marcoulides, G.A. (2000). Generalizability Theory. In H.E.A. Tinsley, & S. Brown (Eds.). *Handbook of Applied Multivariate Statistics and Mathematical Modeling* (p. 527-551). San Diego, CA: Academic Press.

- Moss, P. A. (1992). Shifting Conceptions of Validity in Educational Measurement: Implications for Performance Assessment. *Review of Educational Research*, 62(3), 229-258.
- Mushquash, C., and O'Connor, B. P. (2006). SPSS and SAS Programs for Generalizability Theory Analyses. *Behavior Research Methods*, 36(3), 542-547.
- Nalbantoğlu, F. (2009). *Performans Ölçümlerinde Genellenebilirlik Kuramıyla Farklı Desenlerin Karşılaştırılması*. Yüksek Lisans Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Nie, Y., Yeo, S. M., and Lau, S. (2007). Application of Generalizability Theory in The Investigation of The Quality of Journal Writing in Mathematics. *Studies in Educational Evaluation*, 33, 371-383.
- Nocera, F., Ferlazzo, F., and Borghi, V. (2001). G Theory and The Reliability of Psychophysiological Measures: A Tutorial. *Psychophysiological Research*, 38, 796-806.
- Ødegård, A., Hagtvet, K. A., and 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.
- Palm, T. (2008). Performance Assessment and Authentic Assessment: A Conceptual Analysis of the Literature. *Practical Assessment, Research & Evaluation*, 13(4), 1-11.
- Pedersen, G., Hagtvet, K. A., and Karterud, S. (2007). Generalizability Studies of the Global Assessment of Functioning—Split Version. *Comprehensive Psychiatry*, 48, 88– 94.

- Rentz, J. O. (1987). Generalizability Theory: A Comprehensive Method for Assessing and Improving the Dependability of Marketing Measures. *Journal of Marketing Research*, 24(1), 19-28.
- Sharma, S., and Weathers, D. (2003). Assessing Generalizability of Scales Used in Cross National Research. *International Journal of Research in Marketing*, 20, 287-295.
- Shavelson, J. R., and Webb, N. M. (1981). Generalizability Theory:1973-1980. *British Journal of Mathematical and Statistical Psychology*, 34, 133-166.
- Shavelson, J. R., and Webb, N. M. (1991). *Generalizability Theory: A Primer*. Newbury Park. CA: Sage Publications.
- Suen, H. K., and Lei, P.W. (2007). Classical Versus Generalizability Theory of Measurement. *Educational Measurement*, 4, 1-13.
- Şencan, H. (2005). *Sosyal Davranışsal Ölçümlerde Güvenirlik ve Geçerlilik*. Ankara: Seçkin Yayıncılık.
- 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.
- Touchie, C., Humphrey, M.S., Ainslie, M., Myers, K., and Wood, T.J. (2010). Two Models of Raters in a Structured Oral Examination: Does It Make a Difference? *Advances in Health Sciences Education: Theory and Practice*, 15(1), 97-108.
- Webb, N. M., and Shavelson, R. J. (1981). Multivariate Generalizability of General Educational Development Ratings. *Journal of Educational Measurement*, 18(1), 13-22.

- Webb, N. M., Shavelson, R. J., and Haertel, E. H. (2006). Reliability Coefficients and Generalizability Theory. *Handbook of Statistics, 26*, 81-124.
- Wei, X., and 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., and Chiu, T. (2001). *Scoring Performance Assessments Based on Judgements: Generalizability Theory*. Kluwer Academic Publishers, Boston.
- Yelboğa, A. (2007). *Klasik Test Kuramı ve Genellenabilirlik Kuramına Göre Güvenirliğin Bir İş Performansı Ölçeği Üzerinde İncelenmesi*. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Yelboğa, A. (2008, Mayıs). *Genellenebilirlik Kuramı ile Bir Personel Seçme Yönteminde Güvenirliğin Değerlendirilmesi*. I. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi'nde sunuldu, Ankara.
- Yelboğa, A. ve Tavşancıl, E. (2010). Klasik Test ve Genellenebilirlik Kuramı'na Göre Güvenirliğin Bir İş Performansı Ölçeği Üzerinde İncelenmesi. *Kuram ve Uygulamada Eğitim Bilimleri, 10*(3), 1825-1854.
- Yılmaz Nalbantoğlu, F. ve Gelbal, S. (2011). İletişim Becerileri İstasyonu Örneğinde Genellenebilirlik Kuramı'yla Farklı Desenlerin Karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 41*, 509-518.
- Yin, P. (2005). A Multivariate Generalizability Analysis of the Multistate Bar Examination. *Educational and Psychological Measurement, 65*, 668-686.

Zibrowski, E. M., Myers, K., Norman, G., and Goldszmidt, M. A. (2011). Relying on Others' Reliability: Challenges in Clinical Teaching Assessment. *Teaching and Learning in Medicine*, 23 (1), 21 -27.

EKLER

**Ek 1. Tıp Eğitimi ve Bilişimi Anabilim Dalı İntramuskuler Enjeksiyon
Yapma Becerisi Değerlendirme Formu**

Katılımcının Adı Soyadı		Dönem	1
Fakülte No		Kanal	
Tarih		İstasyon No	4
Gözlemcinin Adı Soyadı		Sıra No	
Değerlendirme Kriterleri			
Evet: Doğru ve yerinde uygulama, Hayır: Yanlış uygulama ya da atlama.			
Açıklamalar			
Katılımcının ellerini yıkadığı ve hastaya işlem hakkında bilgi verdiği, uygulama sonunda katılımcının yeniden ellerini yıkadığı varsayılmaktadır.			
No	Basamaklar	Evet	Hayır
1	Eldiveni giydi.		
2	Enjektöre ilacı çekmeden önce ilacın uygunluğunu kontrol etti.		
3	Enjektörü ambalajından çıkartıp iğnesini steril olarak taktı.		
4	İlacı enjektöre çekti. (1 cc çekilecek)		
5	Enjektörü dik tutup hafifçe vurarak hava kabarcıklarının üst bölgede toplanmasını sağladı.		
6	Enjektörün pistonunu hafifçe iterek havayı çıkarttı.		
7	Enjeksiyon bölgesini alkollü mendille 5 cm çapında içten dışa doğru dairesel hareketlerle bir defada sildi.		
8	Enjeksiyon bölgesini pasif elin 1 ve 2. parmaklarıyla tespit etti.		
9	Enjektörü iğnenin açık ucunu üstten göreceğ şekilde aktif el ile kalem tutar gibi tutarak dik açıyla ve hızla dokuya girdi.		
10	İğne doku içinde iken enjektörü aktif elle kıpırdatmadan sabit tuttu. (El değiştirmeden)		
11	Enjektörü pistonunu pasif elle geri çekerek kas içine girildiğini kontrol etti.		
12	İlacı yavaşça, dokuda basınç oluşturmayacak hızda pasif elle enjekte etti.		
13	Kuru pamukla bastırırken enjektörü aynı açı ve hızla geri çekti.		
14	Kanama duruncaya kadar enjeksiyon alanına kuru pamukla hafifçe bastırdı.		
15	İğnenin koruyucu kapağını kapatmadan enjektörü delici kesici atık kutusuna attı.		
16	Kullanılan malzemeleri kırmızı (tıbbi atık) çöp torbasına attı.		
17	Eldivenleri çıkardı ve çöp torbasına attı.		
Toplam Evet Puanı			
Genel Değerlendirme	Kalır	Sınırdadır	Geçer
			İyi Düzeyde Geçer
			Mükemmel Düzeyde Geçer

Ek 2. Dengelenmiş (ö:p) x g Deseni Syntax ve Analiz Sonuçları

CONTROL CARDS FOR RUN 1
Control Cards File Name: ~Temp.txt

```
GSTUDY   dengelenmis desen
COMMENT
COMMENT   Processing date: 11.12.2011 10:50:57
COMMENT
COMMENT
COMMENT_PATH: C:\Desktop\dengelenmis desen veri.txt
COMMENT
COMMENT%  item   (i)
COMMENT%  rater  (r)
COMMENT%  person (p)
OPTIONS  NREC 5 "*.lis" TIME NOBANNER
EFFECT   * i      17
EFFECT   r        8
EFFECT   p:r      30 30 30 30 30 30 30 30
FORMAT   30 0
PROCESS  "~Temp.dat"
```

ANOVA TABLE FOR RUN 1

Effect	df	T	SS	MS	VC
i	16	8.62403	8.62402	0.53900	0.00141
r	7	3.18824	3.18824	0.45546	0.00043
p:r	232	21.64315	18.45490	0.07955	0.00199
ir	112	34.21569	22.40343	0.20003	0.00514
ip:r	3712	222.34903	169.67843	0.04571	0.04571
Mean		0.00001			
Total	4079		222.34902		
Grand Mean:		0.00004			

EXPECTED VALUES OF T TERMS FOR RUN 1

ET() terms do not include (total number of observations) X
(grand mean squared)

$$\begin{aligned}
 ET(i) &= 17.00*VC(ip:r) + 510.00*VC(ir) + 17.00*VC(p:r) + 510.00*VC(r) \\
 &\quad + 4080.00*VC(i) \\
 ET(r) &= 8.00*VC(ip:r) + 240.00*VC(ir) + 136.00*VC(p:r) + 4080.00*VC(r) \\
 &\quad + 240.00*VC(i) \\
 ET(p:r) &= 240.00*VC(ip:r) + 240.00*VC(ir) + 4080.00*VC(p:r) + 4080.00*VC(r) \\
 &\quad + 240.00*VC(i) \\
 ET(ir) &= 136.00*VC(ip:r) + 4080.00*VC(ir) + 136.00*VC(p:r) + 4080.00*VC(r) \\
 &\quad + 4080.00*VC(i) \\
 ET(ip:r) &= 4080.00*VC(ip:r) + 4080.00*VC(ir) + 4080.00*VC(p:r) + 4080.00*VC(r) \\
 &\quad + 4080.00*VC(i) \\
 ET(Mean) &= 1.00*VC(ip:r) + 30.00*VC(ir) + 17.00*VC(p:r) + 510.00*VC(r) \\
 &\quad + 240.00*VC(i)
 \end{aligned}$$

EXPECTED MEAN SQUARE EQUATIONS FOR RUN 1

$$\begin{aligned}
 EMS(i) &= 1.000*VC(ip:r) + 30.000*VC(ir) + 240.000*VC(i) \\
 EMS(r) &= 1.000*VC(ip:r) + 30.000*VC(ir) + 17.000*VC(p:r) + 510.000*VC(r) \\
 EMS(p:r) &= 1.000*VC(ip:r) + 17.000*VC(p:r) \\
 EMS(ir) &= 1.000*VC(ip:r) + 30.000*VC(ir) \\
 EMS(ip:r) &= 1.000*VC(ip:r)
 \end{aligned}$$

*** EMS matrix is upper diagonal***

Date and time at beginning of Run 1: Sun Dec 11 10:50:57 2011
Processor time for run: 0 seconds

The calculated grand mean = 0,9422
This value has been subtracted from the actual scores for the calculations.

While this improves the accuracy, it does not affect the calculated variances.

Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	6,00
'p'	Random	20,00

```
-----
```

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	6,00	r	Delta only
p:r	0,000	120,00	r	Delta only
i r	0,001	6,00	dr	Delta and delta
i p:r	0,000	120,00	dr	Delta and delta

RESULTS:

s2(T) = 0,001
 s2(D) = 0,001
 s2(d) = 0,001
 Er2 = 0,533
 Phi = 0,515

```
=====
```

 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	6,00
'p'	Random	40,00

```
-----
```

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	6,00	r	Delta only
p:r	0,000	240,00	r	Delta only
i r	0,001	6,00	dr	Delta and delta
i p:r	0,000	240,00	dr	Delta and delta

RESULTS:

s2(T) = 0,001
 s2(D) = 0,001
 s2(d) = 0,001
 Er2 = 0,574
 Phi = 0,556

```
=====
```

 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	6,00
'p'	Random	60,00

```
-----
```

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1,00	d	tau only
r	0,000	6,00	r	Delta only
p:r	0,000	360,00	r	Delta only
i r	0,001	6,00	dr	Delta and delta
i p:r	0,000	360,00	dr	Delta and delta

RESULTS:

s2(T)	= 0,001
s2(D)	= 0,001
s2(d)	= 0,001
Er2	= 0,589
Phi	= 0,571

=====
 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	8,00
'p'	Random	15,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1,00	d	tau only
r	0,000	8,00	r	Delta only
p:r	0,000	120,00	r	Delta only
i r	0,001	8,00	dr	Delta and delta
i p:r	0,000	120,00	dr	Delta and delta

RESULTS:

s2(T)	= 0,001
s2(D)	= 0,001
s2(d)	= 0,001
Er2	= 0,579
Phi	= 0,563

=====
 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	8,00
'p'	Random	30,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	8,00	r	Delta only
p:r	0,000	240,00	r	Delta only
i r	0,001	8,00	dr	Delta and delta
i p:r	0,000	240,00	dr	Delta and delta

RESULTS:

s2(T)	= 0,001
s2(D)	= 0,001
s2(d)	= 0,001
Er2	= 0,629
Phi	= 0,612

=====
 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	8,00
'p'	Random	45,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	8,00	r	Delta only
p:r	0,000	360,00	r	Delta only
i r	0,001	8,00	dr	Delta and delta
i p:r	0,000	360,00	dr	Delta and delta

RESULTS:

s2(T)	= 0,001
s2(D)	= 0,001
s2(d)	= 0,001
Er2	= 0,647
Phi	= 0,630

=====
 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	10,00
'p'	Random	12,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	10,00	r	Delta only

p:r	0,000	120,00	r	Delta only
i r	0,001	10,00	dr	Delta and delta
i p:r	0,000	120,00	dr	Delta and delta

RESULTS:

s2(T)	= 0,001
s2(D)	= 0,001
s2(d)	= 0,001
Er2	= 0,612
Phi	= 0,596

=====

Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	10,00
'p'	Random	24,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	10,00	r	Delta only
p:r	0,000	240,00	r	Delta only
i r	0,001	10,00	dr	Delta and delta
i p:r	0,000	240,00	dr	Delta and delta

RESULTS:

s2(T)	= 0,001
s2(D)	= 0,001
s2(d)	= 0,001
Er2	= 0,667
Phi	= 0,651

=====

Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	10,00
'p'	Random	36,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	10,00	r	Delta only
p:r	0,000	360,00	r	Delta only
i r	0,001	10,00	dr	Delta and delta
i p:r	0,000	360,00	dr	Delta and delta

RESULTS:

s2 (T)	=	0,001
s2 (D)	=	0,001
s2 (d)	=	0,001
Er2	=	0,687
Phi	=	0,672

Ek 3. Dengelenmemiş (ö:p) x g Deseni Syntax ve Analiz Sonuçları

CONTROL CARDS FOR RUN 1
Control Cards File Name: ~Temp.txt

```
GSTUDY   dengelenmemis desen
COMMENT
COMMENT  Processing date: 11.12.2011 09:57:19
COMMENT
COMMENT
COMMENT_PATH: C:\Desktop\dengelenmemis veri.txt
COMMENT
COMMENT% item   (i)
COMMENT% rater  (r)
COMMENT% person (p)
OPTIONS  NREC 5 "*.lis"  ET  EMS SAT TIME NOBANNER
EFFECT   * i           17
EFFECT   r             8
EFFECT   p:r          23 25 26 28 30 31 36 41
FORMAT   30 0
PROCESS  "~Temp.dat"
```

ANOVA TABLE FOR RUN 1

Effect	df	T	SS	MS	VC
i	16	8.90638	8.90637	0.55665	0.00142
r	7	2.82274	2.82273	0.40325	0.00032
p:r	232	21.35074	18.52800	0.07986	0.00193
ir	112	35.08440	23.35529	0.20853	0.00541
ip:r	3712	228.52721	174.91481	0.04712	0.04712
Mean		0.00001			
Total	4079		228.52721		
Grand Mean:		0.00004			

EXPECTED VALUES OF T TERMS FOR RUN 1

ET() terms do not include (total number of observations) X
(grand mean squared)

$$\begin{aligned}
 ET(i) &= 17.00*VC(ip:r) + 527.85*VC(ir) + 17.00*VC(p:r) + 527.85*VC(r) \\
 &\quad + 4080.00*VC(i) \\
 ET(r) &= 8.00*VC(ip:r) + 240.00*VC(ir) + 136.00*VC(p:r) + 4080.00*VC(r) \\
 &\quad + 240.00*VC(i) \\
 ET(p:r) &= 240.00*VC(ip:r) + 240.00*VC(ir) + 4080.00*VC(p:r) + 4080.00*VC(r) \\
 &\quad + 240.00*VC(i) \\
 ET(ir) &= 136.00*VC(ip:r) + 4080.00*VC(ir) + 136.00*VC(p:r) + 4080.00*VC(r) \\
 &\quad + 4080.00*VC(i) \\
 ET(ip:r) &= 4080.00*VC(ip:r) + 4080.00*VC(ir) + 4080.00*VC(p:r) + 4080.00*VC(r) \\
 &\quad + 4080.00*VC(i) \\
 ET(Mean) &= 1.00*VC(ip:r) + 31.05*VC(ir) + 17.00*VC(p:r) + 527.85*VC(r) \\
 &\quad + 240.00*VC(i)
 \end{aligned}$$

EXPECTED MEAN SQUARE EQUATIONS FOR RUN 1

$$\begin{aligned}
 EMS(i) &= 1.000*VC(ip:r) + 31.050*VC(ir) + 240.000*VC(i) \\
 EMS(r) &= 1.000*VC(ip:r) + 29.850*VC(ir) + 17.000*VC(p:r) + 507.450*VC(r) \\
 EMS(p:r) &= 1.000*VC(ip:r) + 17.000*VC(p:r) \\
 EMS(ir) &= 1.000*VC(ip:r) + 29.850*VC(ir) \\
 EMS(ip:r) &= 1.000*VC(ip:r)
 \end{aligned}$$

*** EMS matrix is upper diagonal***

Date and time at beginning of Run 1: Sun Dec 11 09:57:19 2011
Processor time for run: 0 seconds

The calculated grand mean = 0,9404

This value has been subtracted from the actual scores for the calculations.

While this improves the accuracy, it does not affect the calculated variances.

Facets with adjusted sample sizes according to Brennan

'i'	Differentiation	17,00
'r'	Random	7,73
'p'	Random	30,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	7,73	r	Delta only
p:r	0,000	240,00	r	Delta only
i r	0,001	7,73	dr	Delta and delta
i p:r	0,000	240,00	dr	Delta and delta

RESULTS:

s2(T) = 0,001
s2(D) = 0,001
s2(d) = 0,001
Er2 = 0,613
Phi = 0,600

=====
Facets with suggested sample sizes

'i' Differentiation 17,00
'r' Random 5,92
'p' Random 20,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	5,92	r	Delta only
p:r	0,000	118,40	r	Delta only
i r	0,001	5,92	dr	Delta and delta
i p:r	0,000	118,40	dr	Delta and delta

RESULTS:

s2(T) = 0,001
s2(D) = 0,001
s2(d) = 0,001
Er2 = 0,520
Phi = 0,507

=====
Facets with suggested sample sizes

'i' Differentiation 17,00
'r' Random 7,73
'p' Random 15,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	7,73	r	Delta only
p:r	0,000	115,95	r	Delta only
i r	0,001	7,73	dr	Delta and delta
i p:r	0,000	115,95	dr	Delta and delta

RESULTS:

s2(T) = 0,001
s2(D) = 0,001
s2(d) = 0,001
Er2 = 0,562

Phi = 0,549

=====
Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	9,55
'p'	Random	12,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	9,55	r	Delta only
p:r	0,000	114,60	r	Delta only
i r	0,001	9,55	dr	Delta and delta
i p:r	0,000	114,60	dr	Delta and delta

RESULTS:

s2(T)	= 0,001
s2(D)	= 0,001
s2(d)	= 0,001
Er2	= 0,592
Phi	= 0,580

=====
Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	5,95
'p'	Random	40,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	5,95	r	Delta only
p:r	0,000	238,00	r	Delta only
i r	0,001	5,95	dr	Delta and delta
i p:r	0,000	238,00	dr	Delta and delta

RESULTS:

s2(T)	= 0,001
s2(D)	= 0,001
s2(d)	= 0,001
Er2	= 0,562
Phi	= 0,548

=====
 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	7,96
'p'	Random	30,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	7,96	r	Delta only
p:r	0,000	238,80	r	Delta only
i r	0,001	7,96	dr	Delta and delta
i p:r	0,000	238,80	dr	Delta and delta

RESULTS:

s2(T)	= 0,001
s2(D)	= 0,001
s2(d)	= 0,001
Er2	= 0,618
Phi	= 0,605

=====
 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	9,56
'p'	Random	24,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	9,56	r	Delta only
p:r	0,000	229,44	r	Delta only
i r	0,001	9,56	dr	Delta and delta
i p:r	0,000	229,44	dr	Delta and delta

RESULTS:

s2(T)	= 0,001
s2(D)	= 0,001
s2(d)	= 0,001
Er2	= 0,648
Phi	= 0,636

=====
 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	5,98
'p'	Random	60,00

```
-----
```

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	5,98	r	Delta only
p:r	0,000	358,80	r	Delta only
i r	0,001	5,98	dr	Delta and delta
i p:r	0,000	358,80	dr	Delta and delta

RESULTS:

s2(T) = 0,001
 s2(D) = 0,001
 s2(d) = 0,001
 Er2 = 0,578
 Phi = 0,565

```
=====
```

 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	7,94
'p'	Random	45,00

```
-----
```

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	7,94	r	Delta only
p:r	0,000	357,30	r	Delta only
i r	0,001	7,94	dr	Delta and delta
i p:r	0,000	357,30	dr	Delta and delta

RESULTS:

s2(T) = 0,001
 s2(D) = 0,001
 s2(d) = 0,001
 Er2 = 0,636
 Phi = 0,623

```
=====
```

 Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	7,98
'p'	Random	45,00

```
-----
```

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	7,98	r	Delta only
p:r	0,000	359,10	r	Delta only
i r	0,001	7,98	dr	Delta and delta
i p:r	0,000	359,10	dr	Delta and delta

RESULTS:

s2(T) = 0,001
 s2(D) = 0,001
 s2(d) = 0,001
 Er2 = 0,637
 Phi = 0,624

=====

Facets with suggested sample sizes

'i'	Differentiation	17,00
'r'	Random	9,36
'p'	Random	36,00

Pattern	Var. Comp.	Levels	Signature	Rule
i	0,001	1.00	d	tau only
r	0,000	9,36	r	Delta only
p:r	0,000	336,96	r	Delta only
i r	0,001	9,36	dr	Delta and delta
i p:r	0,000	336,96	dr	Delta and delta

RESULTS:

s2(T) = 0,001
 s2(D) = 0,001
 s2(d) = 0,001
 Er2 = 0,664
 Phi = 0,652