

**GENELLENEBİLİRLİK KURAMINDA ÇOK YÜZEYLİ
DESENLERİN İNCELENMESİ**

YÜKSEK LİSANS TEZİ

ZARİFE TAŞTAN

**MERSİN ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME ve DEĞERLENDİRME BİLİM DALI**

**MERSİN
TEMMUZ-2017**

GENELLENEBİLİRLİK KURAMINDA ÇOK YÜZEYLİ DESENLERİN İNCELENMESİ

YÜKSEK LİSANS TEZİ

ZARİFE TAŞTAN

MERSİN ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

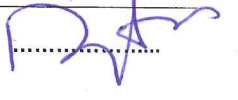


EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME ve DEĞERLENDİRME BİLİM DALI

Danışman

Yrd. Doç. Dr. N. Bilge UZUN

MERSİN
TEMMUZ-2017

Zarife TAŞTAN tarafından Yrd. Doç. Dr. N. Bilge UZUN danışmanlığında hazırlanan "Genellenebilirlik Kuramında Çok Yüzeyle Desenlerin İncelenmesi" başlıklı bu çalışma aşağıda imzaları bulunan jüri üyeleri tarafından oy birliği ile Yüksek Lisans tezi olarak kabul edilmiştir.

Görevi	Ünvanı, Adı ve Soyadı	İmza
Başkan	Doç. Dr. Devrim ALICI	
Üye	Yrd. Doç. Dr. Gülşen TAŞDELEN TEKER	
Danışman	Yrd. Doç. Dr. N. Bilge UZUN	

Yukarıdaki Jüri kararı Eğitim Bilimleri Enstitüsü Yönetim Kurulu'nun 01/09/2020 tarih ve 26/16 sayılı kararıyla onaylanmıştır.



Bu tezde kullanılan özgün bilgiler, şekil, tablo ve fotoğraflardan kaynak göstermeden alıntı yapmak 5846 sayılı Fikir ve Sanat Eserleri Kanunu hükümlerine tabidir.

ETİK BEYAN

Mersin Üniversitesi Lisansüstü Eğitim-Öğretim Yönetmeliğinde belirtilen kurallara uygun olarak hazırladığım bu tez çalışmada,

- Tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- Görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlâk kurallarına uygun olarak sunduğumu,
- Başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- Atıfta bulunduğum eserlerin tümünü kaynak olarak kullandığımı,
- Kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- Bu tezin herhangi bir bölümünü Mersin Üniversitesi veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı,
- Tezin tüm telif haklarını Mersin Üniversitesi'ne devrettiğimi

beyan ederim.

ETHIC DECLARATION

This thesis is prepared in accordance with the rules specified in Mersin University Graduate Education Regulation and I declare to comply with the following conditions:

- I have obtained all the information and the documents of the thesis in accordance with the academic rules.
- I presented all the visual, auditory and written informations and results in accordance with scientific ethics.
- I refer in accordance with the norms of scientific works about the case of exploitation of others' works.
- I used all of the referred works as the references.
- I did not do any tampering in the used data.
- I did not present any part of this thesis as an another thesis at Mersin University or another university.
- I transfer all copyrights of this thesis to the Mersin University.

03 Temmuz 2017/03 July 2017

Zarife TAŞTAN



ÖZET

GENELLENEBİLİRLİK KURAMINDA ÇOK YÜZEYLİ DESENLERİN İNCELENMESİ

Araştırmanın temel amacı; Genellenebilirlik kuramına göre geometri performans görevlerini puanlamada öğrencilerin birden fazla puanlayıcı tarafından puanlanmasıyla oluşturulan desenlerden elde edilen G çalışmaları sonuçlarını incelemektir.

Araştırmanın çalışma grubunu, 2014- 2015 öğretim yılında Mersin ili merkez liselerinde öğrenim gören 128 12. sınıf öğrencisi oluşturmaktadır. Ayrıca araştırmada öğrencilerin geometri performans görevlerinin puanlanmasında sekiz puanlayıcı görev almıştır. Araştırmada puanlayıcıların öğrencileri sekiz görev doğrultusunda puanlamasıyla oluşturulan (b:o)xgxp , (b:o)x(g:p) ve (b:(p:o))xg desenleri (b: öğrenci, g: görev, p: puanlayıcı, o: okul) için ayrı ayrı G çalışması yapılmıştır.

Elde edilen sonuçlara göre, aynı veri seti üzerinden kurgulanan desenlerde varyans bileşenlerine ilişkin büyüklükleri, varyans değişimlerinin büyüklük açısından sıralamadaki yerleri farklılıklar göstermektedir. Bu durumun, söz konusu bileşenlere ilişkin yorumlamaları da değiştirdiği belirtilmiştir. (b:o)xpxg ve (b:o)x(g:p) desenlerinde okula ilişkin varyans değişiminin diğer bileşenlere göre küçük olması okullardan kaynaklı bir farklılığın çok az biçiminde yorumlanırken; (b:(p:o))xg deseninde okullara ait varyans bileşeninin çok yüksek olması, okullarda geometri becerileri bakımından farklılaşmanın fazla olduğu biçiminde yorumlanmıştır. Bunun yanı sıra söz konusu desenlerdeki artık varyans değeri de değişmektedir. (b:o)xpxg deseni için kestirilen artık varyans bileşeni, (b:o)x(g:p) ve (b:(p:o))xg desenlerine göre oldukça küçük bulunmuştur. Bu bağlamda G kuramında duruma ilişkin kurgulanan desenin doğruluğunun yapılacak yorumların güvenilirliğini ve geçerliğini etkileyeceği sonucuna ulaşılmıştır. Her bir desen için elde edilen G ve Phi katsayıları dikkate alındığında (b:o)xpxg desenine ait G ve Phi katsayılarının (b:o)x(g:p) desenine göre yüksek olmasının sebebi; artık varyans değerinin (b:o)x(g:p) desenine göre oldukça düşük olması; (b:o)xpxg deseninin G ve Phi katsayılarının (b:(p:o))xg desenine göre düşük olması ise varyans bileşenlerinin sayıca fazla olması ile açıklanmıştır.

Bu çıkarımlar doğrultusunda (b:(p:o))xg deseninin varyans bileşenlerini açıklama oranı daha az gibi görünse de G ve Phi katsayılarının diğer iki desene göre daha yüksek çıktığı sonucuna ulaşılmıştır. Bunun nedeni de desenin yapısındaki varyans bileşen sayısının az olması olarak açıklanmıştır.

Anahtar Sözcükler: klasik test kuramı, genellenebilirlik kuramı, çok yüzeysel desen, performans ölçme, puanlama anahtarı

Danışman: Yrd.Doç.Dr. N.Bilge Uzun, Mersin Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme Ve Değerlendirme Bilim Dalı, Mersin.

ABSTRACT

INVESTIGATION OF MULTI-SURFACE PATTERNS IN GENERALIZABILITY

The main purpose of the research; in scoring geometry performance tasks according to the generalizability theory, is to examine the results of the G tasks obtained from the patterns generated by the scores of the students by multiple raters.

The study group of the study constitutes 128 12th grade students studying at high school in the center of Mersin province during 2014-2015 school year. In addition, eight raters took part in the scoring of students' geometry performance. Tasks in the survey in which the students of the raters were scored in eight task scales in (b: o) xgxp, (b: o) x (g:p) and (b:(p:o))xg (b: student, g: task, p: scorer, o: school) patterns.

According to the obtained results, it is seen that the sizes of the variance components in the patterns plotted on the same dataset show differences in the order of the variance changes in size. It is also indicated that this situation changed the interpretation of the subject components. While it is interpreted that there is no school-based difference as the difference of variance in (b:o)xpxg and (b: o) x (g: p) is smaller compared to the other components, (b: (p: o)) xg pattern, the variance component of the schools is too high, and the differences in terms of the geometric skills of the schools are interpreted as being too high. In addition to this, the residual variance in the patterns in question also changes. The residual variance component predicted for (b: o) xpxg pattern is quite small compared with the (b: o) x (g: p) and (b: (p: o)) xg patterns. In this context, the correctness of the pattern of the situation in G theory has been reached as the result that the interpretations to be made will affect the reliability and validity. Considering the G and Phi coefficients obtained for each pattern, the reason why the G and Phi coefficients of the (b: o) xpxg pattern are higher than the pattern of (b: o) x (g: p) is that the residual variance value is significantly lower than the (b: o) x (g: p) design (b: o) xpxg pattern is lower than the G and Phi coefficients (b: (p: o)) xg is explained that the number of variance components is large.

In the direction of these inferences the variance components of the (b: (p: o)) xg pattern appear to be less descriptive but it is concluded that the G and Phi coefficients are higher than the other two. The reason for this is explained as there are small number of variance components in the pattern structure.

Key Words: classical test theory, generalizability theory, multi-surface pattern, performance measurement, scoring key.

Advisor: Assist. Prof. N. Bilge Uzun, Mersin University, Department of Educational Sciences, Department of Measurement and Evaluation in Education, Mersin.

TEŞEKKÜR/ÖNSÖZ

Çalışmalarımın her aşamasında beni yönlendiren, desteği ve fikirleri ile çalışmamı tamamlamama katkıda bulunan, yorulduğumda bana pozitif enerjisiyle güç veren, benimle birlikte emek sarf eden değerli hocam ve tez danışmanım Yrd. Doç. Dr. N. Bilge UZUN' a

Kıymetli zamanımı benim tez çalışmama ayırıp değerlendiren ve yapıcı önerileriyle çalışmama katkıda bulunan; ayrıca yüksek lisans eğitimim boyunca her konuda benden desteğini esirgemeyen kıymetli hocam Doç. Dr. Devrim ALICI' ya;

Yardımlarını eksik etmeyen, gecenin geç saatinde rahatsız etsemde sorularıma içtenlikle cevap veren, fikir ve önerileriyle yol gösteren, aynı zamanda tez jüri komitesinde yer alan değerli hocam Yrd. Doç. Dr. Gülşen TAŞDELEN TEKER' e;

Mersin Üniversitesinde yüksek lisansa başladığım ilk günden, tez çalışmamı tamamladığım son güne kadar benden desteklerini esirgemeyen, kendimi geliştirmemde bana öncülük eden Ölçme Değerlendirme Anabilim Dalı'ndaki tüm hocalarıma;

Tez çalışmama fikir ve önerileriyle yol gösteren ve değerli katkılarıyla tezimin şekillenmesine yardımcı olan Doç. Dr. Neşe GÜLER'e;

Yüksek lisansa başladığım ilk günden tezimi savunduğum güne kadar samimiyetiyle hep yanımda olan, bana yol gösteren ,yardımlarını esirgemeyen değerli arkadaşım Arş. Gör. Mehtap AKTAŞ' a

Çalışmamı okuyup fikir ve önerileriyle destek veren, başım her sıkıştığında yanımda bulabileceğim, bana moral veren arkadaşım, dostum Ceren MUTLUER' e;

Bana her konuda güvenen, benim bugünlere gelmemde en büyük destekçim olan sevgili ANNEM ve BABAMA;

Arada mesafeler olsa da çalışma sürecinde sürekli arayarak motivasyonumu yüksek tutmamı sağlayan canım ablam Şeriban KAYA' ya;

Tezimi uzatmama neden olsa da, hayatıma kattığı anlam ve tarifsiz sevgi için neşe kaynağım kızım İdil Neva' ya ayrıca her koşulda yanımda olan, maddi manevi desteğini esirgemeyen, benimle birlikte emek veren, hayatımda iyiki var dediğim eşim Erkan TAŞTAN' a çok teşekkür ederim.

İÇİNDEKİLER

	Sayfa
İÇ KAPAK	
ONAY	
ETİK BEYAN	
ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
İÇİNDEKİLER	iv
TABLolar DİZİNİ	vi
ŞEKİLLER DİZİNİ	vii
KISALTMALAR ve SİMGELER	viii
BÖLÜM I	1
1. GİRİŞ	1
1.1. Problem	1
1.1.1. Hata türleri	2
1.1.1.1. Sabit Hatalar	2
1.1.1.2. Sistematik Hatalar	2
1.1.1.3. Tesadüfi (Seçkisiz) Hatalar	3
1.1.2. Hata Kaynakları	3
1.1.2.1. Ölçmeciden Kaynaklanan Hatalar	3
1.1.2.2. Ölçme Aracından Kaynaklanan Hatalar	3
1.1.2.3. Ölçülen Kişiden Kaynaklanan Hatalar	3
1.1.2.4. Ölçme Ortamından Kaynaklanan Hatalar	4
1.1.3. Klasik Test Kuramından Genellenebilirlik Kuramına	5
1.1.4. Genellenebilirlik Kuramı İle İlgili Kavramlar	8
1.1.5. Çaprazlanmış (Crossed) ve Yuvalanmış (Nested) Desen	9
1.1.6. Dengelenmiş ve Dengelenmemiş Desenler	10
1.1.7. Mutlak Hata Varyansı ve Bağlı (Görelî) Hata Varyansı	10
1.1.8. Genellenebilirlik Katsayısı (G-katsayısı) ve Güvenirlik Katsayısı (Phi-katsayısı)	12
1.1.9. Genellenebilirlik (G) Çalışması ve Karar (D) Çalışması	14
1.1.10. Rastgele (Random) ve Sabit (Fixed) Yüzeyler	16
1.1.11. İlgili Araştırmalar	18
1.1.11.1. Yurt İçinde Yapılan Çalışmalar	18
1.1.11.2. Yurt Dışında Yapılan Çalışmalar	25
1.2. Amaç	28
1.3. Problem Cümlesi	29
1.3.1. Alt Problemler	29
1.4. Önem	29
1.5. Sayıtlar	30
1.6. Sınırlılıklar	30
BÖLÜM II	31
2. YÖNTEM	31
2.1. Araştırmanın Modeli	31
2.2. Çalışma Grubu	31
2.3. Veri Toplama Araç ve/veya Teknikleri	31
2.4. Verilerin Analizi	32

	Sayfa
BÖLÜM III	40
3. BULGULAR	40
3.1. Bireyin (b) Okul (o) Yüzeyinde Yuvalandığı, Görev (g) ve Puanlayıcı (p) Yüzeylerinin ise Çapraz Tasarlandığı (b:o)×gxp Deseninin Genellenebilirlik Çalışması Sonuçları	40
3.2. Bireylerin (b) Okullarda (o) ve Görevlerin (g) Puanlayıcılarda (p) Yuvalandığı ve Bu İkisinin Çapraz Tasarlandığı (b:o)×(g:p) Deseninin Genellenebilirlik Çalışması Sonuçları	43
3.3. Puanlayıcının (p), Okul (o) Yüzeyinde Yuvalandığı, Birey Yüzeyinin de Puanlayıcı (p) Yüzeyinde Yuvalandığı ve Görev (g) Yüzeyinin Çapraz Tasarlandığı (b:(p:o))×g Deseninin Genellenebilirlik Çalışması Sonuçları	45
BÖLÜM IV	48
4. SONUÇ, TARTIŞMA VE ÖNERİLER	45
4.1. Öneriler	51
KAYNAKLAR	52
EKLER	54
ÖZGEÇMİŞ	63

TABLolar DİZİNİ

	Sayfa
Tablo 1.1. Klasik Test Kuramı ve Genellenebilirlik Kuramı Arasındaki Temel Ayrımlar	17
Tablo 2.1. (b:o) x g x p Desenine Ait Veri Giriş Tablosu	33
Tablo 2.2. (b:o) x (g:p) Desenine Ait Veri Giriş Tablosu	35
Tablo 2.3. (b:(p:o)) x g Desenine Ait Veri Giriş Tablosu	37
Tablo 3.1. Bireyin (b) Okul (o) Yüzeyinde Yuvalandığı, Görev (g) ve Puanlayıcı (p) Yüzeylerinin ise Çapraz Tasarlandığı (b:o) x g x p Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri	39
Tablo 3.2. Bireyin (b) Okul (o) Yüzeyinde ve Görevin (g) Puanlayıcı (p) Yüzeyinde Yuvalandığı ve Bu İkisinin Çapraz Tasarlandığı (b:o) x (g:p) Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri	43
Tablo 3.3. Puanlayıcının (p), Okul (o) Yüzeyinde Yuvalandığı, Birey Yüzeyinin de Puanlayıcı (p) Yüzeyinde Yuvalandığı ve Görev (g) Yüzeyinin Çapraz Tasarlandığı (b:(p:o)) x g Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri	46

ŞEKİLLER DİZİNİ

	Sayfa
Şekil 1.1. bmxp Desenine Ait Mutlak Hata Varyansının Varyans Bileşenleri	11
Şekil 1.2. bmxp Desenine Ait Bağlı Hata Varyansının Varyans Bileşenleri	12
Şekil 2.1. (b:o)xp Desenine Ait Varyans Kaynakları	34
Şekil 2.2. (b:o)x(g:p) Desenine Ait Varyans Kaynakları	35
Şekil 2.3. (b:(p:o))xg Desenine Ait Varyans Kaynakları	38

KISALTMALAR ve SİMGELER

Kısaltma/Simge	Tanım
G Kuramı :	Genellenebilirlik Kuramı
G Çalışması :	Genellenebilirlik Çalışması
K Çalışması :	Karar Çalışması
G Katsayısı :	Genellenebilirlik katsayısı
KTK :	Klasik Test Kuramı
Phi Katsayısı:	Güvenirlilik Katsayısı



BÖLÜM I

1. GİRİŞ

1.1. Problem

Teknolojinin hızla geliştiği günümüzde bilimsel bilgi de hızla gelişmektedir. Bu gelişim eğitimde de gelişmelere ve değişmelere sebep olmuştur. Bununla birlikte ölçme ve değerlendirme alanında da yeni düzenlemelere gidilmiştir. Daha önceleri öğrenci başarısını sadece sonuca odaklanarak değerlendiren, öğrencinin bilgisini klasik ölçme araçları (kısa yanıtı, uzun yanıtı, çoktan seçmeli sınavlar vb.) ile sınırlı bir zaman diliminde ölçmeye çalışan ölçme ve değerlendirme yaklaşımları bu değişimlerle beraber öğrencinin değerlendirilmesi konusunda farklı bir bakış açısı kazanmıştır. Artık öğrencilerden beklenen; bilgiyi ezberleyen değil, kazandığı bilgiyi farklı yaşam durumlarına uygulayan, problem çözebilen, eleştirel düşünebilen, verileri analiz edebilen ve etkili sunabilen, sözlü ve yazılı iletişim kurabilen, kendini değerlendirebilen bireyler olmalarıdır. Bu nedenle uygulanan sonuç odaklı klasik değerlendirme yaklaşımları yerine, sözü edilen üst düzey bilişsel özelliklerin ölçülmesinde öğrenme sürecinin de değerlendirilmeye katılması önemli görülmektedir. Bu nedenle eğitimde performans değerlendirme daha sık yer almaya başlamıştır.

Performans, "eğitimdeki öğrenme çıktılarını kazanmadaki başarıma" olarak tanımlanır ve bilişsel, duyuşsal, psiko-motor davranışlarla üst düzey zihinsel beceriler olmak üzere bütün öğrenme ürünlerini içerir (Turgut ve Baykul, 2010). Performans geniş anlamda bütün öğrenme çıktıları olarak alındığından, performansın ölçülmesi de, her tür ve düzeydeki öğrenme ürününün ölçülmesidir. Performansa dayalı durum belirleme (performans değerlendirme) öğrencilerin daha üst düzey düşünme gerektiren karmaşık yapıdaki bazı görevleri yerine getirmesini amaç edinen bir değerlendirme türü olup; ölçme ve değerlendirme alanında önemli bir yer tutmaktadır. Performans değerlendirme, öğrencilerin bilgiyi yeniden yapılandırmasını sağladığından, öğrencilerin güçlü ve zayıf yanlarını, başarılarını ve eksikliklerini, öğretim yöntemlerinin yeterliğini anlamada yardımcı olur. Bu amaçla öğrencileri değerlendirmek için klasik ölçme araçlarının (kısa yanıtı, uzun yanıtı, çoktan seçmeli sınavlar vb.) yanı sıra performans değerlendirme araçları kullanılabilir.

Performansa dayalı değerlendirmeler "öğrencilere anlamlı ve ilgi çekici olan görevlerin yerine getirilmesi yoluyla bilgi, beceri ve çalışma alışkanlıklarının uygulanması için bir dizi stratejiyi temsil eder" (Hibbard ve diğerleri, 1996, s.5; akt: Birualdi, 1998). Bu tür bir değerlendirme, öğretmenin bir öğrencinin bilgiyi nasıl anladığını ve uyguladığını anlatmaktadır.

Ayrıca öğretmenler, öğrencilere ek öğrenme deneyimi sağlamak için performansa dayalı değerlendirmeleri öğretim sürecine entegre edebilirler (Birualdi, 1998).

Performans değerlendirmede verilecek karar için ölçmeye konu olan özellik, gözlenen değer yardımıyla elde edilmeye çalışıldığından yapılan ölçme işlemi, eğitimde yapılan diğer ölçme işlemlerinde olduğu gibi hatalardan arınık olamaz. Bu nedenle, performans değerlendirmede de verilecek kararların doğruluğu açısından yapılan ölçme işlemlerinin güvenilir olması oldukça önemlidir. Yapılan ölçme işlemleri sonrası, ölçmeye konu olan özellik bakımından kişiler, nesnelere ya da olaylar hakkında elde edilen ölçme sonuçlarına dayanarak belli kararlar verilmektedir. Dolayısıyla, eğitim öğretim süreci içerisinde ölçme ve değerlendirmenin önemi büyüktür. Verilen kararların doğruluğu değerlendirmenin dayandığı ölçümlerin olabildiğince az hatalı olmasına bağlıdır.

Psikolojik ölçümlerde bireyin tepkileri gözlemlendiğinden, ölçme sonuçlarına mutlaka belli oranda hata karışır. Ölçmeye dahil olan hata türleri ve hata kaynakları aşağıda sıralanmıştır (Atılğan,2009; Erkuş, 2003).

1.1.1. Hata türleri

Ölçme hataları, kaynağı, yönü ve miktarı göz önüne alındığında sabit, sistematik ve tesadüfi (seçkisiz) hatalar olmak üzere üç türde incelenebilir.

1.1.1.1.Sabit Hatalar:

Bir ölçmeden diğerine, miktarı değişmeyen hatalardır. Sabit hatalar, her bir ölçmede aynı miktar ve yöndedirler (Baykul, 2010; Erkuş, 2003). Örnek olarak bir öğretmenin her okuduğu cevap kağıdına 10 puan fazla verdiğini düşünelim. Elde edilen ölçümler hatalı olacaktır. Bu hatalar her bir ölçme için aynı yönde ve aynı miktarda olacaktır.

1.1.1.2. Sistematik Hatalar

Ölçülen büyüklüğe, ölçmeye veya ölçme koşullarına bağlı olarak sistematik olarak değişen değişken hatalardır. Sistematik hatalar, belirli bir yönde ilerleyici yada gerileyici miktarlarda meydana çıkarlar. Örnek olarak öğretmenin okuduğu cevap kağıtlarından yazısı güzel olana 5 puan fazla vermesi sistematik hatadır.

1.1.1.3. Tesadüfi(Seçkisiz) Hatalar:

Ölçme sonuçlarına gelişigüzel karışan, kaynağı, miktarı ve yönü kesin olarak bilinmeyen hatalardır. Öğretmenin cevap kağıdını dikkatsiz okuması ve puanları yanlış toplaması, sınav günü öğrencinin rahatsız olması, şans başarısı gibi etkenler bu tür hataların ortaya çıkmasına neden olurlar.

Sabit ve sistematik hatalar ölçme aracından veya ölçmeyi yapandan kaynaklandığından ve çoğu durumda miktarı belirlenebildiğinden dolayı hata önceden alınacak önlemlerle teorik olarak da olsa düzeltilebilmektedir (Turgut ve Baykul, 2010). Fakat eğitimdeki ölçmelerde hatanın miktarını belirlemek ve düzeltmek çoğu zaman mümkün olmaz. Seçkisiz hatalar kontrol edilemediğinden sabit ve sistematik hatalarda olduğu gibi cebirsel işlemlerle düzeltilemez. Bu nedenle güvenilirliğe (ya da güvenilmezliğe) konu olan hatalar genellikle sistematik olmayan tesadüfi (seçkisiz) hatalardır (Erkuş, 2003).

1.1.2. Hata Kaynakları

1.1.2.1. Ölçmeciden Kaynaklanan Hatalar

Ölçmecinin taraflı davranması, dikkatsiz davranması (yanlış hesaplama) gibi test uygulandıktan sonra ortaya çıkan hatalardır. Ölçmecinin yaşı, psikolojik özellikleri, motivasyonu, bilgi ve eğitim düzeyi ve cinsiyeti hataya yol açabilir. Aynı zamanda uygulayıcının testi alanlara yaptığı gereksiz müdahalelerde stres yaratarak ölçme aracının güvenilirliğini etkileyebilir.

1.1.2.2. Ölçme Aracından Kaynaklanan Hatalar

Soruların anlaşılammaması, baskı hatası, sorunun okunaksız olması, yoruma açık olması ölçme aracından kaynaklanan hatalardır.

1.1.2.3. Ölçülen Kişiden Kaynaklanan Hatalar

Ölçülen bireylerin içinde buldukları özel koşullardan yada olumsuz yaşantılardan dolayı gerçek performanslarını ortaya koyamamaları güvenilirliği etkileyebilir. Ölçülen kişiden kaynaklanan hata olarak, öğrencinin derse tutumu, kişiliği, yorgunluğu, hastalığı, soruyu yanlış okuması ve kaydırma yapması sayılabilir.

1.1.2.4. Ölçme Ortamından Kaynaklanan Hatalar

Ölçme aracı standart bir ortamda uygulanmadığında, bireyler arasında çıkan farklılıkların, gerçekten bireylerin gerçek puanlarındaki farklılıklardan mı kaynaklandığı bilinemez. Sınav yerinin ısı, ışık ve havalandırmasının yetersiz olması, gürültülü bir ortam olması güvenilirliği düşürmede etkilidir.

Performans değerlendirmede güvenilir puanlama sağlamak objektif testler kadar kolay olmamakla birlikte gözleme dayalı bir değerlendirme olduğundan birçok hatayı da içinde barındırmaktadır. Bu bağlamda performans değerlendirmede elde edilen ölçümlerin hatasız, dolayısıyla ona bağlı olarak verilecek kararların olabildiğince doğru olabilmesi açısından, performans ölçme işleminde kullanılan araçların da belli niteliklere sahip olması beklenmektedir. Bu nitelikler; geçerlik, güvenilirlik ve kullanılabilirlik. **Geçerlik**; bir ölçme aracının kullanılış amacına hizmet etme derecesidir. Ölçme aracı, ölçülecek özelliği tam ve doğru olarak ve ölçmeye dahil olmayan başka değişkenlerle karıştırmadan ölçebiliyorsa geçerlidir. Örneğin; sınav sorularının derste işlenen konuları dengeli olarak kapsamaması, kopya çekilmiş olması, puanlamada yanlı davranılması vb. durumlarda sınavdan elde edilecek puanların geçerliği düşük olur. Diğer bir nitelik olan **güvenirlik**, bir ölçme aracının hatalardan arınık olarak ölçme yapabilme yeterliğidir. Güvenirlik katsayısı, ölçmedeki hatasızlığı belirtir. Hata azaldıkça güvenilirlik katsayısı da artacaktır. **Kullanılabilirlik** ise; ölçme aracının ölçtüğü özelliği doğru ve tutarlı bir şekilde kolay ve masrafsız olarak ölçmesidir (Atılgan, 2009; Tekin, 2010; Turgut ve Baykul, 2010).

Performans değerlendirmede öğrencinin performansının puanlayıcılar tarafından değerlendirilmesi söz konusu olduğundan bu çalışmalarda, puanlayıcı güvenilirliği ön plana çıkmaktadır. Puanlayıcı güvenilirliği, birden çok sayıdaki puanlayıcının aynı özelliğe verdikleri puanların tutarlılığıdır. Performans ölçümlerinde, yalnızca puanlayıcı güvenilirliğinin ön plana alınması yorumların güvenilirliğini tehlikeye atabilir. Çünkü; puanlayıcılar performansın ölçülmesinde güvenilirliği etkileyen önemli bir hata kaynağı olmakla birlikte ölçümlere karışan diğer değişkenlik kaynakları ve puanlayıcıların birey, görev gibi birçok değişkenlik kaynaklarıyla etkileşimleri de güvenilirlik için önemlidir (Güler, 2008). Bu nedenle güvenilirlik belirlenirken birçok değişkenlik kaynağından gelen hatalar ve bu değişkenlik kaynaklarının birbirleriyle etkileşimleri sonucu çıkabilecek hatalar da dikkate alınmalıdır.

Performans değerlendirmeye yönelik kullanılan araçlar; dereceli puanlama anahtarları, kontrol listeleri, gözlem formları, öğrenci ürün dosyaları (portfolyo), vb. dir. Performans değerlendirmede sıklıkla puanlayıcı güvenilirliğini arttırmak için çalışmaya uygun olarak hazırlanmış dereceli puanlama anahtarlarının kullanılması önerilir. *Dereceli puanlama* anahtarları öğrencilerin performanslarını ve cevaplarını belirlenen ölçülere göre puanlamada

kullanılan kılavuzlardır (Turgut ve Baykul, 2010). Moskal, Barbara M. ve Jon A. Leydens (2000)'e göre puanlama anahtarları, puanlayıcıların kim olduğuna bakılmaksızın tutarlı puanlama yapmaya yardımcı olmalıdır. Puanlama anahtarları, puanlayıcılar arasındaki farklılıkları tamamen ortadan kaldırmasa da, iyi tasarlanmış bir puanlama anahtarı, puanlayıcılar arasındaki tutarsızlıkları azaltabilir. Puanlama kriterlerinin önceden tanımlanması güvenilirliği artıracaktır. Dereceli puanlama anahtarının bütünsel (holistik) ve çözümlenmeli (analitik) olmak üzere iki türü vardır. Bu çalışmada da kullanılan *bütünsel dereceli puanlama* anahtarı, performansı bir bütün olarak görerek bütüne puan vermeyi sağlarken *analitik puanlama* anahtarı performansın puanlanacak özelliklerinin ayrı ayrı puanlanmasını sağlar. Analitik puanlama anahtarı, öğrenme eksikliklerini saptama ve öğretimi iyileştirme amacıyla da kullanılabilir.

Yapılan bir ölçme işleminde, hatanın kaynağının ne olduğu, ölçmeye hangi koşulların hata karıştırdığını bulmak önemli bir sorundur. Örneğin; ölçmeye konu olan özelliğin zeka olduğu durumda; bu özelliği ortaya çıkarmak için kullanılan bir zeka testinden elde edilen puan, testi alanın zekasını ne derece yansıtmaktadır? Elde edilen puan ölçmeye konu olan özellik dışındaki faktörlerden ne derecede arınıktır? Bu sorulara cevap olarak psikoloji ve eğitimde ölçme hataları ilk olarak Klasik Test Kuramı (KTK) ile ele alınmıştır (Brennan, 1992). Daha sonra KTK'nın sınırlılıklarını ortadan kaldırmak için Genellenebilirlik Kuramı (G Kuramı) ortaya atılmıştır. Aşağıdaki bölümlerde KTK'nın bu sınırlılıklarından yola çıkılarak G kuramına ilişkin temel açıklamalar yer almaktadır.

1.1.3. Klasik Test Kuramından Genellenebilirlik Kuramına

Bilimin temel amaçlarından biri; nesne ya da bireylere ilişkin bir takım bilgiler elde etmek; gözlenen özellikler çerçevesinde onları betimleyebilmektir. Ölçme; bir betimleme işi olup gözlemden öte gitmeyecektir. Bu durumda "bireylere ya da nesnelere ait bir özelliğin gözlenen değeri, o özelliğin gerçek değerini ne kadar yansıtmaktadır?" sorusunu yanıtlamak, yapılan betimleme işleminin ne kadar duyarlı bir biçimde yapıldığı konusunda araştırmacılara bilgi sağlayacaktır (Brennan, 1992). Ancak bütün ölçmeler belli bir derecede hata içerdiğinden, bilim adamları ölçmenin hassasiyetini arttırmak ve daha hatasız ölçmeler yapabilmek için çeşitli yollar aramışlardır.

Bu arayışlar sonucunda eğitimde ve psikolojide yapılan ölçmelerde hatayı ele alan KTK'ye göre; bireyin doğrudan gözlenemeyen özellikleri ölçülürken, ölçmeye karışan hatalardan dolayı gerçek değer doğrudan elde edilemez; gözlenen puanlar yardımıyla kestirilmeye çalışılır (Baykul, 2010). Bir zeka testi uygulanan bireylerin tepkilerinden elde edilen sonuçlar; bizim bireylerin zekasına ilişkin gözlemiş olduğumuz değerlerdir. Hatasız ölçme yapılamayacağından;

yapılan ölçme işlemi esnasında, bu sonuçlara çeşitli hatalar karışır. Bu nedenle elimizdeki ölçümler bireylerin gerçek zeka değerlerine karşılık gelmeyebilir. Bu bağlamda *hata*; ölçmeye konu olan özelliğin gerçek değeri ve gözlenen değeri arasındaki farktır. Diğer bir ifade ile KTK' ye göre, bir bireyin gözlenen puanı, gerçek puan ve hata puanının toplamından oluşmaktadır ve bu durum Eşitlik 1'de gösterilmektedir.

$$X_i = T_i + E_i \quad (1)$$

Burada X_i , i. maddenin gözlenen puanı, T_i , i. maddeye ilişkin gerçek puan ve E_i ise hata puanıdır (Crocker ve Algina, 1986).

Hata türlerinden biri olan seçkisiz hataların ölçme sonuçlarının güvenilirliği ile doğrudan ilişkili olduğu bilinmektedir. KTK' de güvenilirlik, test-tekrar test, paralel formlar ve iç tutarlılık ilişkili olmak üzere birbirinden farklı yöntemler ile hesaplanır ve hata kaynakları tahmin edilir. KTK' ye dayalı bu yöntemlerde, ölçmelerde bir kerede sadece bir hata kaynağı dikkate alınır. Örneğin; test tekrar test yönteminde hesaplanan *kararlılık* anlamındaki güvenilirlikte hata kaynağı zamandır. Paralel formlar yöntemiyle hesaplanan *tutarlılık* anlamındaki güvenilirlikte ise formlar hata kaynağı olarak ele alınır. Diğer yandan *iç tutarlılık* anlamında güvenilirliğin hesaplanmasında kullanılan yöntemler ise maddeleri hata kaynağı olarak ele alır (Atılğan, 2004). Dolayısıyla aynı ölçmeye ilişkin olarak KTK' nin farklı hata kaynaklarını dikkate alan farklı güvenilirlik yöntemleri ile elde edilen güvenilirlik katsayıları da farklı olabilir. Hesaplanan güvenilirlik değerleri arasındaki farklılık oldukça fazla olduğunda, hangi güvenilirlik değerinin göz önünde bulundurulacağına karar vermek çok güç olabilir. Bu nedenle, KTK yaklaşımları performans ölçümlerinin güvenilirliğinin hesaplanmasında yetersiz kalabilir.

KTK' nin sınırlılıklarını ortadan kaldırmak için davranış ve sosyal bilimlerde (psikoloji, eğitim) birçok ölçme sorununu belirleyebilmek için geniş bir kavramsal sisteme ve güçlü istatistiklere sahip olan Genellebilirlik (G) Kuramı; Cronbach, Gleser, Nanda ve Rajaratnam tarafından (1972) önerilmiştir (Brennan, 1992).

G kuramının, KTK' ye göre üstün öne çıkan özelliği hatayı ele alış biçimidir. KTK yalnızca seçkisiz hatalarla ilgilenir, ölçme hatalarını farklı değişkenlik kaynaklarından gelen potansiyel hata kaynaklarına ayırmaz ; oysa hata her zaman seçkisiz olmayabilir (Kim, 1993). G kuramı, sistematik olan ve olmayan hata kaynaklarını ayrı ayrı ele alır ve bunların birbirleriyle etkileşimlerini de belirleyebilir (Brennan, 1992). Ayrıca G kuramı güvenilirlik katsayılarını etkileyen çoklu hata kaynaklarının modellenmesine izin verir (Taylor, 2013). G kuramına dayalı yapılan güvenilirlik çalışmalarında; puanlayıcılardan, diğer değişkenlik kaynaklarından (zaman, test formu, madde, görev, vb.) ve bu değişkenlik kaynakları arası etkileşimden ortaya çıkan hata varyansları bir arada göz önüne alınmaktadır. Bununla birlikte bütün potansiyel değişkenlik

kaynaklarından gelen potansiyel hataları birlikte ve eş zamanlı değerlendirerek kapsamlı tek bir güvenilirlik katsayısı hesaplanmasını sağlamaktadır (Atılğan, 2004). Diğer ifadeyle G kuramı, araştırmacıya puanların potansiyel değişkenlik kaynaklarını tanımlamalarına ve her bir kaynağın katkısının ilişkisini analiz etmelerine olanak sağlamaktadır (Taylor, 2013). KTK' ye göre güvenilirlik hesaplanırken, güvenirlüğün hangi amaç için kullanılacağına göz önüne alınması ve hesaplanmasında buna uygun bir yol seçilmesi gerekir (Baykul, 2000). Fakat G Kuramı güvenilirlik yöntemlerinin tümünü içine alır.

G kuramı, ölçmenin daha az hata ile yapılabilmesi için ölçmenin hangi koşullarda yapılması gerektiği konusunda da bilgi verir. G kuramı, ölçme işlemlerinin tekrar edilmesi ile elde edilecek olan gözlenen puanlar arasındaki tutarsızlıkların kaynağını ve miktarını belirlemeye de olanak vermektedir. Aynı zamanda bu kuram test puanlarının ve puanlayıcılar arası tutarlılığın da bir ölçüsüdür (Brennan, 1992).

G kuramında ayrıca, bağıl ve mutlak değerlendirmeler arasında bir ayrım söz konusudur. Bu nedenle güvenilirlik hesaplanırken hata varyanslarına bağlı olarak (bağıl ve mutlak hata) farklı güvenilirlik katsayıları elde edilir. Bir başka ifade ile G kuramı görünüşte benzer olan bazı güvenilirlik katsayıları arasında ayrım yapmaya yardımcı olur (Fan ve Sun, 2014). Fakat KTK' de bağıl ve mutlak hatalar arasındaki ayrım göz önüne alınmadan güvenilirlik hesaplanmaktadır (Brennan, 2001). KTK sonuçlarına dayalı olarak yapılabilecek tek yorum, bireylerin birbirlerine göre durumlarını karşılaştırmaktır. Yani araştırmacı öğrencinin diğer öğrencilerin ne kadarından başarılı olduğunu belirleyebilir. Ancak uygulayıcının araştırdığı şey (mutlak değerlendirme de olduğu gibi), öğrencinin diğer arkadaşlarından bağımsız olarak performans düzeyinin belirlenmesi olduğunda bu bilgi ancak G kuramı tarafından sağlanmaktadır. Diğer bir deyişle KTK, araştırmacılara yalnızca görelî karar verebilecekleri bilgiyi sağlarken, G kuramı hem görelî hem de mutlak kararlar için yeterli bilgiyi aynı anda sunmaktadır (Güler, Kaya Uyanık ve Taşdelen Teker, 2012).

G kuramı kullanılış amacına göre de KTK' den farklılıklar gösterir. Bu bağlamda; G kuramının amacı, ölçme konusu olan bireyler ya da objelerin gözlenen puanlarının evren puanlarına (KTK' deki gerçek puan kavramına karşılık gelir) doğrulukla genellenmesini sağlamaktır. Cronbach'a göre bir testin güvenilir olup olmamasının dışında, kolayca diğer bir dereceye genellenip genellenemediği de önemli bir sorundur (Atılğan, 2004). Diğer bir deyişle, genelleme evreninde elde edilen bulguların, kabul edilebilir gözlemler evrenine genellenip genellenemeyeceği de önemlidir. G kuramının temeli varyans analizine (ANOVA) dayandığından dolayı, varyans analiziyle toplam varyans desendeki bağımsız değişkenlere bölünür. Böylece ölçme sonuçları farklı varyans kaynaklarına ayrılarak, bireylerin ya da objelerin gözlenen puanlarının evren puanlarına (gerçek puanlarına) genellenebilmesi sağlanır. Bu sebeple genellenebilirlik analizlerinde varyans bileşen tahminleri merkezi öneme sahiptir. Bunlar daha

sonraki sonuçların hepsi için çok önemli bir temel oluşturan yapı taşlarıdır. Varyans bileşeni tahminlerinin doğruluğunun, ölçüm hatası varyanslarının ve genellenebilirlik katsayılarının tahmin edilmesi ve yorumlanması üzerinde doğrudan etkisi vardır (Gao ve Brennan,2001) Ayrıca, G kuramında vurgulanan ANOVA konuları, birçok deneysel tasarım ve ANOVA da baskın olanlardan farklıdır. Özellikle G kuramı, F testleri değil; varyans bileşenleri ve tahminleri üzerine yoğunlaşmaktadır (Brennan, 2011).

Aynı zamanda, G kuramı, bir ölçme değerlendirme işlemi optimize edebilmemize de izin verir (Yin ve Shavelson, 2004). Bir başka ifadeyle, istenilen düzeyde güvenilirlik elde etmek için her bir değişkenlik kaynağının koşul sayısının belirlenmesini sağlar.

1.1.4. Genellenebilirlik Kuramı İle İlgili Kavramlar

Genellenebilirlik kuramına göre evrende bir ya da birden çok varyans kaynağı bulunmaktadır. Varyans kaynakları olan madde (ya da görev), zaman, puanlayıcı gibi ölçmenin benzer durumlar setine, *değişkenlik kaynakları* denir. Bu değişkenlik kaynakları ölçmenin yüzeylerini (facet) oluşturur. *Yüzey*, en basit haliyle benzerlik gösteren ölçme durumlarıdır (Brennan, 1992). Bu yüzeylerin de *koşulları* (conditions) vardır. Örneğin, bir başarı testinde maddeler yüzey olarak adlandırılırken testte bulunan her bir madde de bu yüzeyin koşulu olarak düşünülebilir. Yüzeyler varyans analizinde faktörlere, yüzeylerin koşulları ise bir faktörün düzeylerine benzetilebilir (Güler, Kaya Uyanık ve Taşdelen Teker, 2012). Genellikle, varolan bir yüzeyin olası koşullarının sonsuz büyüklükte olduğu varsayılır (Brennan, 1992). *Ölçme objesi*, istenilen kararların alınacağı ölçmenin hedefleridir. Pek çok ölçme durumunda bireyler ölçme objesi durumundadır. G kuramında *evren (universe)* terimi yüzeyler; *popülasyon (population)* terimi ölçme objesi için kullanılır. *Evren puanı*, genelleme evrenindeki bütün koşulların ölçme objeleri için ortalama puanıdır. Bütün ölçme objelerinin evren puanlarının varyansı, *evren puan varyansı* olarak adlandırılır ve bu da KTK' deki gerçek puan varyansına benzemektedir (Brennan ,1992; Shavelson ve Webb, 1991).

Genellenebilirlik (G) çalışması belirli koşullar altında elde edilen ölçümlere dayanır. Bu koşullar genellikle daha geniş setteki koşulların bir temsilcisi durumundadır. G çalışmasında yer alan yüzeylere bağlı koşullardan elde edilen ölçme sonuçlarına *genelleme evreni* denir. Bir başka deyişle, genelleme evreni; araştırmacının genellemek istediği koşulların setidir. Alınabilecek tüm koşullardan elde edilen ölçme sonuçları evrenine ise *kabul edilebilir gözlemler evreni* denir. Örneğin, farklı iki okulda bulunan 12. sınıf öğrencilerine sekiz geometri performans görevinin uygulandığı, öğrencilerin performans görevlerini sekiz farklı puanlayıcının puanladığı durumda üç yüzeyli (okul, performans görevi, puanlayıcı) bir ölçme durumu oluşturulmuştur. Olası tüm sekiz puanlayıcı, sekiz madde ve iki okuldan (çalışma ile aynı sayıda yüzey ve koşullara sahip)

oluşan tüm durumlar genelleme evrenini ve benzer durumlar altında olası tüm puanlayıcı, okul ve maddeleri içeren evren de kabul edilebilir gözlemler evrenini oluşturmaktadır (Güler, Kaya Uyanık ve Taşdelen Teker, 2012).

Burada dikkat edilmesi gereken nokta, bir çalışmanın kabul edilebilir gözlemler evreni o çalışmada yer almayan farklı yüzeyleri içermediğinden, yapılacak Karar (D study) çalışmaları da ancak çalışmada yer alan yüzeylere bağlı olarak kurulabilir; G çalışmasında yer almayan farklı yüzeyler üzerinden K çalışması kurmak mümkün değildir (Güler, Kaya Uyanık ve Taşdelen Teker, 2012).

1.1.5. Çaprazlanmış (Crossed) ve Yuvalanmış (Nested) Desen

G kuramında, değişkenlik kaynakları çapraz (crossed) ya da yuvalanmış (nested) şekilde olabilir. Bir değişkenlik kaynağının koşullarının başka bir değişkenlik kaynağının koşullarıyla örtüştüğü duruma *çapraz desen* denilmektedir. Çapraz olarak tasarlanmış bir desende değişkenlik kaynakları arasında "x" işareti kullanılır. Örneğin; bir sınıftaki öğrencilerin (ö) tamamına 12 performans görevi (g) uygulansın. Bu ölçme sürecinde görev alan 3 puanlayıcı (p) da tüm öğrencilerin tüm görevlerini puanlasın. Bu durumda, öğrenciler aynı görevler doğrultusunda aynı puanlayıcılar tarafından puanlandıkları için öğrenci, görev ve puanlayıcı değişkenleri çaprazlanarak $ö \times g \times p$ desenini oluşturur (ö: öğrenci, g: görev, p: puanlayıcı).

Değişkenlik kaynaklarının diğer değişkenlik kaynaklarıyla örtüşmediği yani çaprazlanmadığı durumlar da söz konusudur. Bir değişkenlik kaynağının koşulları diğer değişkenin bazı koşullarıyla örtüşüyorsa değişkenlik kaynakları yuvalanmış olarak tasarlanmalıdır. Yuvalanmış olarak tasarlanmış bir desende değişkenler arasında ":" işareti kullanılır. Örneğin; bir sınıftan 20 farklı öğrencinin (ö) her birine, 8'er farklı performans görevi (g) verilsin ve 4 farklı puanlayıcının her biri (p) her öğrencinin 2'şer görevini puanlasın, bu desenlere de tümüyle yuvalanmış desen adı verilir. Bu desende puanlayıcılar görevlerle, görevler de öğrencilerle yuvalanmıştır ve bu desen $p : g : ö$ olarak ifade edilir.

Benzer şekilde bazı değişkenlik kaynaklarının çaprazlanmış bazılarının ise yuvalanmış olduğu desenlerde düzenlenebilir. Örneğin; 3 okul (o) ve okulların her birinden 20 şer öğrenci (ö) seçilsin. Seçilen öğrencilere 12 performans görevi (g) uygulansın. Bu ölçme sürecinde görev alan 3 puanlayıcı (p) da tüm öğrencilerin tüm görevlerini puanlasın. Her okuldan 20 öğrenci seçildiği için öğrenciler okulların içine yuvalanmıştır (ö:o). Her birey bütün performans görevlerini cevapladığı için ve her performans görevi bütün puanlayıcılar tarafından puanlandığı için okul ve öğrenciler, puanlayıcı ve görev yüzeyleri ile çaprazlanmıştır. Bu desen (ö:o)xp₃g şeklinde gösterilir.

Cronbach (1972) ında önerdiği gibi tüm olası değişkenlik kaynaklarına ilişkin hatanın kestirimine imkân verdiğiinden dolayı G çalışmalarında tümüyle çapraz desenler daha çok tercih edilmektedir. Ancak yuvalanmış desenin bazı varyans bileşenlerinin hesaplanmasında serbestlik derecesini artırmasından dolayı yararlı olacağı da ifade edilmektedir (Alkan, 2013).

1.1.6. Dengelenmiş ve Dengelenmemiş Desenler

Genellenebilirlik ve Karar çalışması yapabilmek için oluşturulan desenler araştırma verisine göre dengelenmiş (balanced) ya da dengelenmemiş (unbalanced) olabilir. Yuvalanmış bir yüzeyin koşullarına ait gözlem sayısı tüm yüzey boyunca aynı ise veri dengelenmiş, yüzeyin koşullarına ait gözlem sayısı farklılaşıyorsa veri dengelenmemiş olmaktadır (Brennan, 2001; Shavelson ve Webb, 1991). Örneğin; farklı okullardan öğrencilerin araştırılmaya dahil edildiği, öğrenci (ö) ve okul (o) yüzeylerinin yuvalandığı (ö:o) bir durumda araştırmada kullanılan her bir okuldaki öğrenci sayısı aynı ise veri dengelenmiş, her bir okuldaki öğrenci sayısı farklı ise veri dengelenmemiş olmaktadır. Bir başka örnek olarak, her bir puanlayıcı için puanladığı cevap kağıtlarının sayısı aynı ise dengelenmiş desen; her bir puanlayıcının puanladığı cevap kağıdı farklı sayıda ise desen dengelenmemiştir.

G kuramında, Genellenebilirlik katsayısı (G) ve Phi katsayısı olmak üzere iki tür güvenilirlik katsayısı hesaplanabilmektedir. Fakat hesaplanan bu güvenilirlik katsayıları hata varyansına göre değişiklik göstermektedir. Bu hata varyansları mutlak (absolute) hata ve bağıl (relative) hata varyansı olarak adlandırılmaktadır.

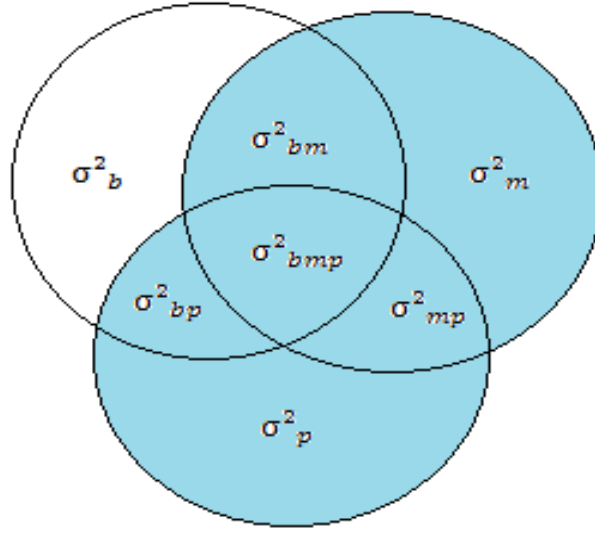
1.1.7. Mutlak Hata Varyansı ve Bağıl (Görelî) Hata Varyansı

Mutlak hata varyansı, bir kişinin gözlenen puanı ve onun evren puanı arasındaki farkın varyansı olarak tanımlanır. Mutlak hata varyansı $\sigma^2(\Delta)$ şeklinde gösterilir ve ölçme objesine ilişkin varyans olan $\sigma^2(b)$ hariç tüm varyans bileşenlerinin toplamını içerir. Bireylerdeki değişim gruba bağlı olmadan mutlak bir ölçüte göre belirlenir.

Örneğin : b x m x p desenli bir çalışmada mutlak hata varyansı (Brennan, 2001; Shavelson ve Webb, 1991).;

$$\sigma^2(\Delta) = \frac{\sigma_m^2}{n_m} + \frac{\sigma_p^2}{n_p} + \frac{\sigma_{bm}^2}{n_m} + \frac{\sigma_{bp}^2}{n_p} + \frac{\sigma_{mp}^2}{n_m n_p} + \frac{\sigma_{bmp}^2}{n_m n_p} \quad (2)$$

şeklinde hesaplanır. bmxp desenine ait mutlak hata varyansının bileşenleri Şekil 1.1.' de yer alan venn şemasında gösterilmiştir.

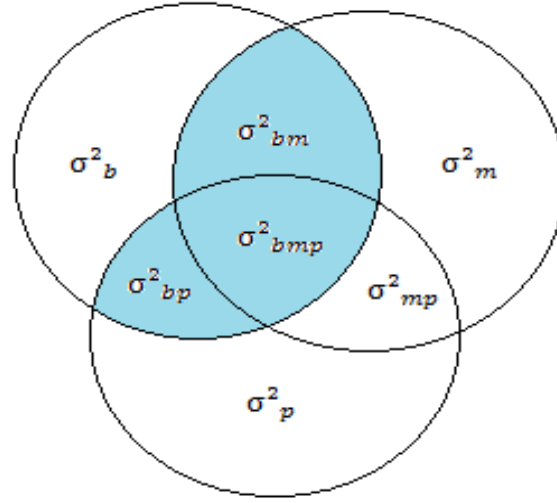


Şekil 1.1. bmxpx Desenine Ait Mutlak Hata Varyansının Bileşenleri

Bağlı (Görel) Hata Varyansı ise; bireylerin gözlenen ve evren puanları arasındaki farklılığın büyüklüğüdür. Gözlenen ve evren puanları popülasyon ortalamalarına bağlı olarak hesaplandığı için “*görel*” ismini alır (Güler, Kaya Uyanık ve Taşdelen Teker, 2012). Bağlı hata varyansı $\sigma^2(\delta)$ şekilde gösterilir ve KTK ’deki hata varyansına karşılık gelir (Brennan, 1992). Bağlı hata varyansında bireylerin başarıları içinde bulunduğu gruptaki puanların sıralamasına bağlıdır (Rentz,1987; akt: Nalbantoğlu, 2009). Bağlı hata varyansı ölçme objesini içeren etkileşimli varyans bileşenlerinin toplamıyla hesaplanır. Örneğin, bmxpx desenli bir çalışmada bağlı hata varyansı (Brennan, 2001; Shavelson ve Webb, 1991);

$$\sigma^2_{(\delta)} = \frac{\sigma^2_{bm}}{n_m} + \frac{\sigma^2_{bp}}{n_p} + \frac{\sigma^2_{bmp}}{n_m n_p} \quad (3)$$

şeklinde hesaplanır. bmxpx desenine ait bağlı hata varyansının bileşenleri Şekil 1.2’ de yer alan venn şemasında gösterilmiştir.



Şekil 1.2. bmxp Desenine Ait Bağlı Hata Varyansının Bileşenleri

G kuramı, bir tek güvenilirlik katsayısı yerine G katsayılarından bahseder ve evren tanımının değişmesiyle katsayının değerinin de değişeceğini öngörür. Bir başka deyişle, G kuramı hem bireylerin performanslarına dayalı bağlı kararlar hem de bireylerin performanslarıyla ilgili mutlak kararlar alınmasına ilişkin iki farklı güvenilirlik katsayısının hesaplanmasına olanak tanır. Bağlı değerlendirmeler için genellenebilirlik katsayısı (G), mutlak değerlendirmeler için ise güvenilirlik katsayısı (Phi) hesaplanır.

1.1.8. Genellenebilirlik Katsayısı (G-katsayısı) ve Güvenirlik Katsayısı (Phi-katsayısı)

G kuramında, hangi modelin uygun olduğu daima uygulama amacına yani araştırmacının ölçme sonucunu nasıl kullanacağına bağlıdır. Sosyal bilimlerdeki ölçmeler; genel olarak bireyleri ya da grupları sıralamak veya bireylerin bilgi, beceri, tutum ve yetenek düzeyini hesaplamak için kullanılmaktadır. Eğer, bireyler birbiriyle kıyaslanacaksa, her bir bireyin bir göreve ilişkin göstermiş olduğu performansın puanlanması, içinde bulunduğu gruptaki diğer bireylerin nasıl performans gösterdiklerine bağlı olarak yapılacaktır (bu bireyin grup içindeki sıralaması nedir?). Bu durumda araştırmacı verilerin analizi için bağlı (göreceli) modeli kullanacaktır. Bağlı hata varyansı kullanılarak hesaplanan Genellenebilirlik (G) katsayısı, bireylerin ham puanlarının büyüklüğündeki olası değişimlere bakılmaksızın, değişkenlik kaynaklarının sıralamadaki yerlerini koruma derecelerini göstermektedir. G katsayısı, KTK' deki güvenilirlik katsayısına karşılık gelmektedir (Mushquash ve O'Connor, 2006).

Araştırmacının bireyleri değerlendirirken mutlak bir ölçüte bağlı olarak değerlendirme yapması durumunda bireylerin performansı, içinde buldukları grubun performansı dikkate

alınmaksızın, önceden belirlenmiş bir ölçüte göre değerlendirilir. Bu durumda araştırmacı mutlak modeli kullanır. Mutlak hata varyansı kullanılarak elde edilen güvenilirlik katsayısı (Phi) , G katsayısından daha katıdır ve ölçülen bireylerin sıralamadaki yerlerinin düzenini yansıtmamasının yanı sıra ham puanlarının büyüklüklerindeki uyumun derecesini de yansıtır. Elde edilen puanların gerçek değerlerinin önemli ya da daha anlamlı olduğu durumlarda Phi katsayısı, araştırmacı için kullanışlı olabilecek bir katsayıdır (Mushquash ve O'Connor, 2006).

KTK' deki gerçek puan G kuramında evren puanı olarak adlandırılır. G katsayısı; evren puan varyansının, evren puan varyansı ile bağıl hata varyansının toplamına oranı ile elde edilir. Phi katsayısı; evren puanı varyansının, evren puanı varyansı ile mutlak hata varyansının toplamına oranı ile elde edilir. Belirli bir kesme puanının üzerindeki puanın önem taşıdığı performans ölçümlerinde Phi katsayısı tercih edilebilir (Brennan, 1992). Elde edilen puanların puan sıralamasındaki yerinin önem taşıdığı durumlarda G katsayısını kullanmak uygun olacaktır. Bağıl ve mutlak kararlar için hesaplanan G katsayılarındaki karışıklığı ortadan kaldırmak üzere, bağıl kararlar için hesaplanan değere G katsayısı, mutlak kararlar için hesaplanan değere Phi katsayısı ya da güvenilirlik (dependability) katsayısı adı verilmektedir.

KTK' de;

$$\text{Güvenirlik katsayısı} = \frac{\text{Gerçek Puan Varyansı}}{\text{Gerçek Puan Varyansı} + \text{Hata Varyansı}} \quad (4)$$

G kuramında;

$$\text{Genellenebilirlik katsayısı} = \frac{\text{Evren puan varyansı}}{\text{Evren Puan Varyansı} + \text{Bağıl Hata Varyansı}} \quad (5)$$

$$\text{Phi Katsayısı} = \frac{\text{Evren puan varyansı}}{\text{Evren Puan Varyansı} + \text{Mutlak Hata Varyansı}} \quad (6)$$

olarak ifade edilir (Brennan, 2001; Shavelson ve Webb, 1991).

Eşitlik 5 ve 6'da belirtildiği üzere, G ve Phi katsayısı sırasıyla bağıl hata ve mutlak hata varyansına bağlı olarak farklı amaçlar doğrultusunda kullanılmaktadır. Aynı desen için evren puan varyansları aynı olacağından; katsayıları belirleyen, hata varyanslarıdır. Mutlak hata varyansı; yapısı gereği bağıl hata varyansına göre daha çok varyans bileşeni içerdiğinden; Phi katsayısı, G katsayısından küçük olma eğilimindedir. Sonuç olarak hata varyans bileşeni arttıkça güvenilirlik katsayıları küçülecektir. Tek yüzeyli, çaprazlanmış desende G katsayısı, Cronbach alfa katsayısının aynısıdır (Sudweeks, Reeve ve Bradshaw, akt: Alharby, 2006).

G kuramıyla, Genellenebilirlik çalışması (G) ve Karar çalışması (D) olmak üzere iki tür çalışma yapılabilir. Aşağıda G ve K çalışmalarına ilişkin açıklamalar bulunmaktadır.

1.1.9. Genellenebilirlik (G) Çalışması ve Karar (D) Çalışması

Genellenebilirlik (G) çalışmasında varyans analizi modeli, değişkenlik kaynaklarının özelliklerine ve sayısına bağlı olarak belirlenir. G çalışması düzenlemede ilk olarak potansiyel değişkenlik kaynaklarının (madde ya da görev, puanlayıcı, zaman vb.) ve ölçme örnekleminin genelleneceği evrenin belirlenmesi gereklidir. Başka bir ifade ile G çalışmaları mümkün olduğunca en kapsamlı şekilde, kabul edilebilir gözlemler evrenini tanımlamalıdır (Shavelson ve Webb, 1991).

G çalışması sadece mevcut ölçmelerin teknik niteliklerini değerlendirmek için değil, gelecekteki ölçme desenleri için değerlendirme sonuçlarını geliştirmede kullanılır (Taylor, 2013). Diğer bir ifadeyle; mevcut ölçme için kullanılan desenden elde edilen varyans bileşenleri, güvenilirliği yükseltmek için farklı bir desen kullanıldığında nasıl değişimlenebilir sorusuna cevap bulunabilir. Daha önce de belirtildiği gibi G çalışmalarında puan çeşitliliğinin (varyans bileşenlerinin) bütün kaynakları ve etkileşimleri ANOVA kullanılarak tahmin edilir ve örneklem evrene genellenir. Bu tahmin edilen varyans bileşenleri ise bir sonraki çalışma olan Karar (K) çalışmasında kullanılır. K çalışması "...olursa ne olur ?" sorusuna cevap arar. Bu aşamada araştırmacı güvenilirliği yükseltmek ve hatayı azaltmak için birçok farklı senaryo dener. Örneğin, "Daha fazla puanlayıcı kullanırsam ne olur? Madde sayısını azaltırsam ne olur?"... gibi. Yapılan bir K çalışmasında, incelenen bireyleri tanımlamak için veri toplanabilir. Bireyler arasında seçme yapmak ya da bireyleri bir programa yerleştirmek, bir sınavdaki grupları karşılaştırmak ya da iki ya da daha fazla değişken arasındaki ilişkiyi araştırmak için K çalışması düzenlenebilir (Alharby, 2006). K çalışmaları, ölçme işlemlerinin daha güvenilir olabilmesi için yüzeylerin farklı sayıdaki koşulları ele alındığında varyans bileşenlerinin tahmin edilebilmesini ve yorumlanmasını sağlar (Brennan, 1992). Sonuç olarak, G çalışmasında elde edilen varyans tahminleri üzerine bir çok K çalışması yapılandırılabilir.

Bu durumda önemli bir ayrıntı da; G çalışmasında kullanılan yüzeylerin koşul sayılarının K çalışmasında kullanılanlarla aynı olmak zorunda olmamasıdır (Brennan, 1992). Benzer biçimde K çalışması için örneklem sayısının G çalışmasının örneklem sayısı ile aynı olması da gerekmemektedir (Chien, 2008).

Örnek olarak, bütün öğrencilerin (b) testte yer alan bütün maddeleri (m) cevaplandığı ve iki farklı puanlayıcı (p) tarafından tüm öğrencilerin tüm cevaplarının puanlandığı durumda, öğrenciler, maddeler ve puanlayıcıların tümüyle çaprazlandığı bir desen (b x m x p) elde edilir. G kuramında, madde ve puanlayıcı değişkeni gibi iki yüzeyin yer aldığı ve yüzeylerin çapraz olduğu bir desene ait gözlenen puan aşağıdaki biçimde tanımlanır (Brennan,1992; Shavelson ve Webb, 1991):

$$X_{bmp} =$$

μ	genel ortalama
$+\mu_b - \mu$	birey etkisi
$+\mu_m - \mu$	madde etkisi
$+\mu_p - \mu$	puanlayıcı etkisi
$+\mu_{bm} - \mu_b - \mu_m + \mu$	birey x madde etkisi
$+\mu_{bp} - \mu_b - \mu_p + \mu$	birey x puanlayıcı etkisi
$+\mu_{mp} - \mu_m - \mu_p + \mu$	madde x puanlayıcı etkisi
$+\mu_{bmp} - \mu_{bm} - \mu_{bp} - \mu_{mp} + \mu_b + \mu_m + \mu_p - \mu$	artık etkisi

Bu durumda, toplam gözlenen varyans aşağıda ifade edilen yedi bağımsız varyans bileşeninden oluşur (Alharby, 2006):

$$\sigma^2(X_{bmp}) = \sigma^2(b) + \sigma^2(m) + \sigma^2(p) + \sigma^2(bm) + \sigma^2(bp) + \sigma^2(mp) + \sigma^2(bmp) \quad (7)$$

K-Çalışmasına örnek olarak iki maddenin ve iki puanlayıcının bulunduğu, çaprazlanmış desen ve bağıl model düzenlenmiş olsun. Bu durumda bağıl varyans (Güler, 2008):

$$\sigma^2(\delta) = \frac{\sigma_{bm}^2}{n_m} + \frac{\sigma_{bp}^2}{n_p} + \frac{\sigma_{bmp}^2}{n_m n_p} = \frac{\sigma_{bm}^2}{2} + \frac{\sigma_{bp}^2}{2} + \frac{\sigma_{bmp}^2}{2 \times 2} \quad (8)$$

olacaktır. Burada σ_b^2 gerçek varyanstır ve σ_m^2, σ_p^2 ve σ_{mp}^2 ise etkileşimsiz varyanslar olarak düşünülürler ve model bağıl olduğu için eşitlikte yer almazlar. Bu durumdaki G-katsayısı, gerçek varyansın, gerçek varyans ile etkileşimli varyansların toplamına bölünmesiyle elde edilir (Güler, 2008):

$$G - \text{katsayısı} = \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_{bm}^2}{2} + \frac{\sigma_{bp}^2}{2} + \frac{\sigma_{bmp}^2}{2 \times 2}} \quad (9)$$

Aynı durum için bağıl model yerine mutlak model kullanıldığı düşünüldüğünde, puanlayıcı, görev ve bunlar arasındaki etkileşime bağlı varyans bileşenleri, mutlak varyansın parçaları olarak ele alınır ve önceki hesaplamada yer alan paydaya eklenerek daha büyük bir

payda elde edilir. Bu durum elde edilecek katsayının küçülmesine sebep olur. Böylece, güvenilirlik katsayısı denilen Phi katsayısı;

$$\begin{aligned}\Phi - \text{katsayısı} &= \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_m^2}{n_m} + \frac{\sigma_p^2}{n_p} + \frac{\sigma_{bm}^2}{n_m} + \frac{\sigma_{bp}^2}{n_p} + \frac{\sigma_{mp}^2}{n_m n_p} + \frac{\sigma_{bmp}^2}{n_m n_p}} \\ &= \frac{\sigma_b^2}{\sigma_b^2 + \frac{\sigma_m^2}{2} + \frac{\sigma_p^2}{2} + \frac{\sigma_{bm}^2}{2} + \frac{\sigma_{bp}^2}{2} + \frac{\sigma_{mp}^2}{2 \times 2} + \frac{\sigma_{bmp}^2}{2 \times 2}}\end{aligned}\quad (10)$$

şeklinde elde edilir.

Bu eşitliklerden, değerlendirmeye daha fazla sayıda puanlayıcı ve madde eklemenin, genellenebilirlik kuramı çerçevesinde güvenilirlik katsayısını arttıracığı ve hatayı düşüreceği görülmektedir (Güler, 2008). Performans değerlendirme işlemi en az hata ile gerçekleştirmek, elde edilen sonuçların güvenilirliğini yükseltmenin en önemli yoludur. Bilindiği üzere, performans değerlendirmede birden fazla puanlayıcının kullanılması puanlama güvenilirliğini olumlu yönde etkilemektedir.

1.1.10. Rastgele (Random) ve Sabit (Fixed) Yüzeyler

Bir yüzeyin rastgele ya da sabit olarak ele alınması tamamen araştırmacının kararına bağlıdır. Örneklem büyüklüğü evren büyüklüğünden çok küçükse ve örneklem evrenden tesadüfi olarak seçilmiş ise bu tür örneklemeler *rastgele* olarak adlandırılırlar. Bir diğer ifade ile, koşulları örneklemin ötesinde evrene genellemek istiyorsak yüzeyimiz rastgeledir. Örneğin, bireylerin bir geometri sınavında üçgende alan konusunu ölçmeye yönelik hazırlanmış 8 maddeye vermiş oldukları cevapların dört ayrı puanlayıcı tarafından puanlandığı bir araştırma deseni ele alınsın. Araştırmacının amacı, bu araştırmada yer alan 8 maddeyi üçgende alan konusu ile ilgili tüm maddelere genellemek ise bu durumda madde yüzeyi *rastgele* olacaktır.

Diğer taraftan, eğer araştırmacının amacı örneklemin ötesinde bir genelleme yapmak değilse, ele alınan değişkenlik kaynağı *sabit* olarak tanımlanır. Örneğin, yukarıdaki örnekte araştırmacının amacı geometri sınavında yer alan sadece 8 maddeye ilişkin öğrencilerin düzeylerini belirlemek ise, bu araştırma deseninde yer alan madde yüzeyi *sabit* olacaktır.

Çalışmamızda yer alan iki okul yüzeyi evrendeki olası tüm okullara genellenmek istendiğinden, okul yüzeyi *rastgele* olarak alınmıştır. Ancak sadece bu iki okul ile ilgileniyor ve okul evrenine genellemek gibi bir amacımız bulunmuyor olsaydı bu durumda okul yüzeyi *sabit* olarak tanımlanacaktır.

Bu temel açıklamalar etrafında ilgili alanyazın da dikkate alındığında (Atılğan, 2004; Shavelson ve Webb, 1991; Brennan, 1992; Kim, 1993; Güler, Kaya Uyanık ve Taşdelen Teker, 2012; Taylor, 2013) KTK ve G kuramının karşılaştırıldığı özet bir tablo aşağıda sunulmuştur.

Tablo 1.1: KTK ve G Kuramı Arasındaki Temel Ayrımlar

KTK	G KURAMI
-KTK' ye dayalı geleneksel güvenilirlik yöntemi, ölçmelerde bir kerede sadece bir hata kaynağını dikkate alır.	G kuramı puanlayıcı, zaman, test formu, madde ve görev gibi bütün potansiyel değişkenlik kaynaklarından gelen potansiyel hataları birlikte ve eş zamanlı değerlendirerek kapsamlı tek bir güvenilirlik katsayısı hesaplanmasını sağlar.
KTK, ölçme hatalarını tesadüfi hata olarak ele alır ve farklı değişkenlik kaynaklarından gelen potansiyel hata kaynaklarını ayırmaz.	G kuramı hem sistematik hemde sistematik olmayan hata kaynaklarını ayrı ayrı ele alır ve bunların etkileşimini de belirleyebilir.
KTK araştırmacılara yalnızca görelî karar verebilecekleri bilgi sağlar.	G kuramı hem görelî hem de mutlak kararlar için yeterli bilgiyi aynı anda sunmaktadır.
KTK' de bağıl ve mutlak hatalar arasındaki ayrım göz önüne alınmadan güvenilirlik hesaplanmaktadır.	G kuramında bağıl ve mutlak değerlendirmeler arasında bir ayrım söz konusudur. Bu nedenle güvenilirlik hesaplanırken hata varyansları arasında bağıl ve mutlak hatalara göre farklılık vardır.
KTK' ye göre güvenilirlik hesaplanırken, güvenilirliğin hangi amaç için kullanılacağı göz önüne alınmalı ve hesaplanmasında buna uygun bir yol seçilmesi gerekir.	Genellenebilirlik kuramı güvenilirlik yöntemlerinin tümünü içine alır.
KTK güvenilirlik çalışmalarında sadece ölçmedeki hatalara kaynaklık eden değişkenlik kaynaklarını ele alır.	G Kuramı, güvenilirlik çalışmalarında sadece ölçmedeki hatalara kaynaklık eden değişkenlik kaynaklarını değil, aynı zamanda bu değişkenlik kaynaklarının ortak etkilerini de ele alır.
KTK de sadece bir değişkenlik kaynağının sayısının değiştirilmesinin, güvenilirliği nasıl etkileyeceği Spearman –Brown formülü ile hesaplanabilir.	K çalışmaları; çok sayıda hata kaynağının analizi ile en uygun güvenilirliğe ulaşılması için, değişkenlik kaynaklarının koşullarının sayısının belirlenmesine olanak sağlar. Böylelikle istenilen düzeyde güvenilirlik için her bir değişkenlik kaynağının koşul sayısının belirlenmesi sağlanır.
KTK güvenilirlik ve geçerliği ayrı ayrı ele alır.	G Kuramı geçerlik ve güvenilirlik arasındaki geleneksel farklılığı da bir ölçüde ortadan kaldırmaktadır.

1.1.11. İlgili Araştırmalar

1.1.11.1. Yurt İçinde Yapılan Çalışmalar

Atılğan (2004), iki yıl ardi ardına yapılan müzik öğretmenliği seçme sınavının verilerini kullanarak Genellenebilirlik Kuramı ve Çok Değişken Kaynaklı Rasch Modeli analizleri uygulamıştır. Analizler sonucunda elde edilen verilerle; Genellenebilirlik kuramının tek değişkenli ve çok değişkenli modellerinin istatistikleri karşılaştırılmış ve Genellenebilirlik kuramının alternatif Karar çalışmalarının gerçek durumla tutarlılıkları incelenmiştir. Ayrıca Genellenebilirlik kuramı ve Çok Değişken Kaynaklı Rasch Modeli istatistikleri karşılaştırılmıştır. Araştırma sonuçlara göre, hem Genellenebilirlik Kuramı hem de Çok Değişken Kaynaklı Rasch Modeli yaklaşımlarının her ikisinin de oldukça kullanışlı sonuçlar ürettiği görülmekle birlikte, Genellenebilirlik Kuramı ve Çok değişken Kaynaklı Rasch Modeli ile değişkenlik kaynakları için kestirilen sonuçlarının kısmen tutarlı sonuçlar verdiği belirlenmiştir.

Atılğan (2005), yaptığı bir başka çalışmada, Genellenebilirlik kuramını tanımlamış, kuramla ilgili temel kavramları açıklamış, Klasik Test kuramına göre üstün olan yönlerini vurgulamış ve uygulanmasını da hipotetik olarak ömeklendirmiştir. G-Kuramına göre ölçmenin güvenilirliğinin belirlenmesi amacıyla puanlayıcıların sürece katıldığı hipotetik bir ölçmeyi farklı ve çok hata kaynaklı bir ölçme durumu olarak ele almış ve ölçüt dayanaklı ölçmeler için Phi, norm dayanaklı ölçmeler için G katsayılarını elde ederek bu katsayıların kullanımını göstermiştir. Araştırmada Genellenebilirlik kuramının birçok ölçme deseni için ölçmenin psikometrik özelliklerinin belirlenmesinde ve ölçme aracının geliştirilmesinde KTK yerine kullanılmasının daha uygun olabileceği sonucuna varılmıştır.

Atılğan ve Tezbaşaran (2005), aynı değişkenlik kaynaklarının farklı düzeyleri için, alternatif K çalışmaları senaryolarıyla elde edilen G ve Phi katsayıları ile aynı değişkenlik kaynaklarının aynı düzeyleriyle gerçekte uygulanması sonucunda elde edilen G ve Phi katsayılarının tutarlılıklarını karşılaştırmışlardır. Araştırmada kullanılan özel yetenek sınavlarının alt boyutlardan oluşması nedeniyle, birey (b), görev (g) ve puanlayıcı (p) değişkenlik kaynağı (facets) olmak üzere, G-kuramının çok değişkenli ($bxgxp$) deseni kullanılmıştır. Sonuç olarak; puanlayıcı sayısının artırılması senaryosu ile kestirilen G ve Phi katsayılarının gerçekte olduğundan daha büyük, puanlayıcı sayısının azaltılması yoluyla kestirilen G ve Phi katsayılarının ise gerçekte olduğundan daha küçük çıktığı görülmüştür.

Yelboğa (2007), çalışmasında 2005 ve 2006 yıllarında, iş performansı ölçęi kullanarak elde ettiği gerçek verilerin Klasik Test kuramı ve Genellenebilirlik kuramına göre güvenilirliklerini kestirmiş ve puanlayıcılar arası tutarlılıkları belirlemeye çalışmıştır. Klasik Test kuramında; test tekrar-test ve Cronbach alfa güvenilirlik katsayılarını, puanlayıcılar arası

güvenirlilik için Kendall'ın uyum katsayısı ve genellenebilirlik kuramında ise çok değişkenlik kaynaklı modellerde G ve Phi katsayılarını hesaplamıştır. Araştırmadan elde edilen sonuçlara göre, Klasik Test Kuramında ve Genellenebilirlik kuramının çok değişkenlik kaynaklı modeliyle elde edilen güvenirlilik katsayılarının birbiriyle uyumlu sonuçlar verdiği görülmüştür. İş performansı ölçeğiyle birden fazla puanlayıcıyla puanlama yapılan durumlarda Genellenebilirlik kuramının kullanılmasının uygun olacağı belirtilmiştir.

Güler (2008), KTK, Genellenebilirlik kuramı ve Rasch Modeli üzerine yaptığı çalışmada matematik başarısının ölçülmesinden elde edilen puanların güvenirliliklerini hesaplamış ve üç kuramdan elde edilen sonuçları karşılaştırmıştır. Klasik test kuramında Cronbach alfa güvenirlilik katsayısı, puanlayıcılar arası uyumun belirlenmesinde Kendall'ın konkordans katsayısı, puanlayıcılar arası korelasyon katsayısı ve puanlayıcıların verdikleri puanların ortalamaları arası fark olup olmadığı F testi ile araştırılmıştır. Genellenebilirlik kuramında, bxgxp tümüyle çaprazlanmış desen kullanılarak Genellenebilirlik ve güvenirlilik katsayıları hesaplanmıştır. Çok değişkenlik kaynaklı Rasch ölçme modeli ile birey, puanlayıcı ve madde boyutlarına ilişkin ayrı ayrı güvenirlilik hesaplamaları yapılmıştır. Araştırma bulgularına göre, matematik başarısını ölçmek için kullanılan ölçme aracının güvenilir sonuçlar verdiği, puanlayıcıların birbiriyle uyumlu puanlamalar yaptığı belirlenmiştir. Ayrıca çalışmada puanların güvenirliliğinin belirlenmesinde, elde edilen puanların kullanılma amacına bağlı olarak üç kuramdan hangisinin kullanılması gerektiğinin belirlenmesi ve en az iki kuramdan yararlanılması gerektiği tespit edilmiştir.

Yelboğa (2008), otomotiv sektöründe personel seçme sürecinde yapılandırılmış mülakat formu kullanarak 18 aday mühendisi; her biri otomotiv firmasında farklı birimlerden sorumlu yedi kişilik bir komisyon tarafından altı görev üzerinden birbirinden bağımsız olarak 1 -10 puanları arasında puanlandırmıştır. G kuramında (b x g x p) çaprazlanmış desininin kullanıldığı çalışmada G ve Phi katsayılarını hesaplamış ve alternatif karar çalışması yapılmıştır. Araştırma bulgularına göre, G katsayısının 0,80'den büyük ve yeterli olduğu; ancak Phi katsayısının 0,80'den küçük olduğu bulunmuştur. Yapılan karar çalışmasında G ve Phi katsayılarını artırmak için puanlayıcı sayısının artırılmasından ziyade madde sayısının artırılması gerektiği belirtilmiştir.

Taşdelen (2009), Nedelsky ve Angoff standart belirleme yöntemleri ile elde edilen kesme puanlarını Genellenebilirlik Kuramını kullanarak karşılaştırmıştır. Araştırmanın birinci aşamasında her bir yöntem için farklı hesaplanan ve farklı uzman görüşlerine dayanan kesme puanları kullanılarak yöntemler karşılaştırılmıştır. Araştırmanın ikinci aşamasında ise aynı veriler ile genellenebilirlik çalışması yürütülmüş ve genellenebilirlik katsayıları ve varyans bileşenleri elde edilmiştir. Ayrıca bu yöntemlerin uygulanması için gereken en uygun puanlayıcı sayısı da belirlenmeye çalışılmıştır. Her iki standart belirleme yönteminin birlikte ele alındığı ve tümüyle çaprazlanmış $m \times p \times y$ modeli uygulanarak yapılan G çalışmasında her iki yöntemde de,

maddeler arası farklılıkların ortaya çıkarılabildiği ve puanlayıcılar arasında bir tutarlılığın olduğu görülmüştür. Çalışmadan elde edilen G ve Phi katsayılarına bakıldığında ise Angoff ve Nedelsky yöntemleri ile standart belirleme çalışılırken 10 puanlayıcının yeterli olacağı sonucuna varılmıştır.

Nalbantoğlu (2009), bir performans sınavı olan OSCE sınavında öğrencilerin birden fazla puanlayıcı tarafından birlikte ve dönüşümlü olarak puanlanmasıyla oluşturulan desenlerden elde edilen G ve K çalışmaları sonuçlarını karşılaştırmıştır. Araştırma sonuçlarına göre, incelenen çaprazlanmış ve yuvalanmış desenlere ilişkin kestirilen varyans bileşenlerinin birbiri ile paralellik gösterdiği görülmüştür. Ayrıca puanlayıcıların öğrencileri puanlamada farklılık göstermediği ve puanlayıcılar arası tutarlık sağlandığında öğrencilerin puanlayıcıların hepsi tarafından tek tek puanlanması yerine puanlayıcıların belli sayıdaki öğrencileri dönüşümlü olarak puanlamasının yeterli olduğu sonucuna ulaşılmıştır.

Güler ve Gelbal (2010), matematik başarısının ölçülmesiyle elde edilen puanların güvenilirliğini belirlemede KTK ve Genellenebilirlik kuramından yararlanmışlardır. KTK' ya göre elde edilen puanların iç tutarlılığının 0.92 gibi oldukça yüksek bir değer olduğu görülmüştür. Puanlayıcılar arası uyum ise Kendall'ın uyum (konkordans) katsayısı ile belirlenmiş ve 0.52 olarak bulunmuştur. Genellenebilirlik kuramına göre matematik başarısının ölçülmesiyle elde edilen puanların genellenebilirlik katsayısı 0.92 ve güvenilirlik katsayısı 0.90 bulunmuştur. Puanlayıcı değişkenlik kaynağının toplam varyansı açıklama oranı % 2,1 olup, oldukça düşük bulunmuştur. Elde edilen bulgular, kullanılan ölçme aracının, öğrencilerin matematik başarısını belirlemede güvenilir sonuçlar verdiğini göstermektedir. Çalışmada yer alan dört puanlayıcının puan ortalamaları arasında fark olmakla birlikte, birbirleriyle uyumlu puanlama yaptıkları belirlenmiştir.

Öztürk (2011), voleybol becerilerine ilişkin dereceleme ölçeğinden elde edilen ölçmelerin; Klasik Test Kuramı ve Genellenebilirlik Kuramı ile elde edilen güvenilirlik katsayılarını karşılaştırmıştır. Genellenebilirlik Kuramında bxgxp deseni sonuçları incelendiğinde en yüksek varyans bileşenin gxp ortak etkisi için kestirilen varyans bileşeni olduğu görülmüştür. Güvenirlik katsayıları iç ölçütlere göre incelendiğinde G, Phi katsayısı, Kendall uyum katsayıları ve Cronbach Alfa katsayılarının beklenenden düşük olduğu belirlenmiştir. Analizler neticesinde Klasik Test Kuramı'nda ve Genellenebilirlik Kuramı'nın çok değişkenli modellerde elde edilen güvenilirlik katsayılarının birbirleriyle uyumlu olduğu görülmüştür. Bu araştırma sonucunda, voleybol becerilerine ilişkin dereceleme ölçeğinin güvenilir bir ölçek olarak değerlendirilemeyeceği ifade edilmiştir.

Güler (2011)'in, çalışmasında, 125 öğrencinin 18 maddeye verdiği cevapları 4 farklı puanlayıcı puanlamıştır. Değişkenlik kaynağının maddeler olduğu (bxm) çapraz desen için hesaplanan G katsayısı ile Cronbach alfa değerleri her puanlayıcı için ayrı ayrı hesaplanmış ve

çok düşük değerler elde edilmiştir. Değişkenlik kaynağının maddeler ve puanlayıcılar olduğu tümüyle çapraz desen (bxm_{xp}) için Genellenebilirlik Kuramına dayalı G- Katsayısı ve Phi katsayısı sırasıyla 0.457 ve 0.456 olarak bulunmuştur.

Özberk (2012), araştırmasında, iki farklı varyans bileşeni belirleme yöntemine göre kestirilen standart hata değerleri ile güvenilirlik ve geçerlik üzerinde karşılaştırmalar yapmıştır. Veriler simülasyon ile tek değişkenlik kaynaklı desene göre üretilmiştir. İlk aşamada b_{xm} (b:birey, m:madde) desenine uygun birey-madde matrisi, 60x5 şeklinde veri seti simülasyon ile R yazılımı kullanılarak üretilmiştir. İkinci aşamada b_{xm} desenine uygun olarak değişkenlik kaynakları yazılımında 1000 kere yeniden örneklenmiştir. Standart hatalar, varyans bileşenleri, mutlak ve bağıl hatalar ANOVA ve bootstrap yöntemleri kullanılarak kestirilmiştir. Araştırmada, *boot-m* prosedürünün geçerlik hakkında daha fazla bilgi verdiği, *boot-b* prosedürünün de G Kuramı çalışmalarında evren puanlarını belirlemede daha kesin kestirimler yaptığı sonucuna varılmıştır.

Büyükkıdık (2012), sıradışı (rutin olmayan) matematik problemi çözme becerilerinin değerlendirilmesine yönelik hazırlanan iki adet performans görevinin, analitik ve bütünsel dereceli puanlama anahtarıyla, dört puanlayıcı tarafından puanlanmasından elde edilen puanlara KTK ve Genellenebilirlik kuramı ile puanlayıcılar arası güvenilirlik incelemesi yapmıştır. KTK' de puanlayıcılar arası güvenilirlik analizi, sınıf içi ilişki katsayısı ve puanlayıcılar arası ilişki katsayısı ile gerçekleştirilmiştir. Aynı performanslara, aynı puanlayıcıların analitik ve bütünsel puanlama anahtarı kullanarak verdikleri puanların ortalamaları arasında fark olup olmadığı bağımlı gruplar t testi ile araştırılmıştır. Genellenebilirlik kuramında ise her iki puanlama anahtarı için b_{xgxp} ve anahtar çeşidinin değişkenlik kaynağı olarak alındığı b_{xgxp_a} tümüyle çaprazlanmış desenler kullanılarak güvenilirlik analizi yapılmıştır. Araştırma sonucunda Genellenebilirlik kuramından elde edilen katsayıların KTK' den elde edilen katsayılara göre göreceli olarak daha yüksek olduğu, Genellenebilirlik kuramının uygulamasının KTK' ye göre daha detaylı bilgi sağladığı ve her iki kuramdan elde edilen bulguların da verilerin güvenilirliğinin yüksek olduğunu desteklediği görülmüştür. Genellenebilirlik kuramında kullanılan desenlerin güvenilirliği etkilediği, değişkenlik kaynağı sayısı arttıkça bireyin, toplam varyansı açıklama yüzdesinin azaldığı görülmüştür. Dereceli puanlama anahtarı türünün de ölçümlerin güvenilirliğini etkileyebileceği, bu araştırmada analitik ve bütünsel dereceli puanlama anahtarlarından elde edilen puanların benzer güvenilirlik katsayıları verdiği; ancak analitik dereceli puanlama anahtarından elde edilen puanların göreceli olarak daha yüksek güvenilirliğe sahip olduğu belirlenmiştir. Sonuç olarak, puanlamanın amacına göre puanlama anahtarı ve kuram seçimi yapılması önerilmiştir.

Anıl ve Büyükkıdık (2012), çalışmalarında 6, 7 ve 8. sınıfta öğrenim görmekte olan 132 öğrencinin problem çözme becerisine yönelik hazırlanan 2 performans görevinde sergiledikleri performanslarının 4 puanlayıcı tarafından puanlanması ile elde edilen veriler Genellebilirlik kuramında incelenmiştir. Ölçmenin nesnesi olan farklı sınıflardaki bireyler (b:s); görev, ölçüt ve puanlayıcı yüzeyleri ile çaprazlanarak (b:s)×g×öxp deseni oluşturulmuştur. Araştırma sonunda G çalışması ile kestirilen en büyük varyans bileşeninin farklı sınıflardaki bireyler olduğu görülmüştür. Bu bulgu problem çözme becerisi açısından grubun heterojen olması ile açıklanmıştır. Ayrıca çalışmada sınıf düzeyine ait varyans bileşeninin önemli olmadığı görülmüştür. Bu durumun, üst sınıfların güdülenme düzeyinin alt sınıflardan daha düşük olmasından ya da uygulanan sınıflarda üst düzey düşünme becerilerini sergileyen benzer oranda öğrencinin varlığından kaynaklanabileceği ifade edilmiştir. Yapılan karar çalışmaları ile ölçümlerin güvenilirliğini arttırmada, puanlayıcı sayısının, görev sayısına göre daha etkili olduğu sonucuna ulaşılmıştır.

Aktaş (2013), aynı performans görevlerinin farklı sayıda puanlayıcı tarafından kontrol listesi, dereceleme ölçeği ve analitik rubrik yardımıyla puanlanması durumunda, puanlayıcılar arası güvenilirlikleri Genellebilirlik kuramı çerçevesinde incelemiştir. Bu amaç doğrultusunda, 5. sınıf düzeyindeki öğrencilerin hikâye yazma becerilerini puanlamak amacıyla, kontrol listesi, dereceleme ölçeği ve analitik rubrik hazırlanmıştır. İlköğretim 5. sınıf öğrencilerine yazdırılan hikâyeler arasından seçilen 6 hikâye 45 puanlayıcıya üç farklı puanlama anahtarı ile 10-15 gün aralıklarla puanlatılmıştır. Araştırmaya katılan 45 puanlayıcı içerisinde 2, 3, 5 ve 10 puanlayıcı 100'er örneklem çekilmiştir. Elde edilen 400 örneklem için Genellebilirlik Kuramı'na göre puanlayıcılar arası güvenilirlikler ve elde edilen 1200 hesaplamanın her bir durum için elde edilen 100 örnekleme için ortancaları ve standart hataları hesaplanmıştır. Genellebilirlik Kuramı'na dayalı güvenilirlik kestirimlerinin ortanca değerleri incelendiğinde, 5 puanlayıcının kontrol listesi kullanarak yaptıkları puanlamaların güvenilirliklerinin ortanca değeri hariç olmak üzere, puanlayıcı sayısı ve aynı zamanda kullanılan ölçeğin kategori sayısı arttıkça ortanca değerlerinin de arttığı gözlenmiştir. Genellebilirlik kuramından elde edilen standart hataların, puanlayıcı sayısı arttıkça azaldığı gözlenmiştir. En düşük standart hata değerlerinin, 10 puanlayıcı olması durumunda elde edildiği saptanmıştır. Puanlayıcı sayısı 5 ve kategori sayısı 2 olduğunda, Genellebilirlik kuramında güvenilirlik kestiriminin en yüksek değeri verdiği belirlenmiştir.

Alkan (2013), PISA 2009 Okuma Becerileri performansını değerlendirmede kullanılan açık uçlu soruların, birden fazla puanlayıcı tarafından birlikte ve dönüşümlü olarak puanlanmasıyla elde edilen farklı desenleri G kuramına göre karşılaştırmıştır. Araştırmada, iki farklı senaryo üzerinden G kuramında kullanılmak üzere iki desen tasarlanmıştır. Bu desenlerden ilki, öğrenci (ö), soru (s) ve puanlayıcı (p) değişkenleri olmak üzere, öğrencilerin

aynı beceriler konusunda puanlayıcıların her biri tarafından puanlandığı öxsxp çapraz desendir. İkinci desen ise, her bir puanlayıcının öğrencilerden sadece bir kısmını puanlamasıyla oluşan, öğrenci ve puanlayıcı değişkenlerinin yuvalanmış olduğu, soruların ise bu değişkenlerle çaprazlanmış olduğu (ö:p)xs desendir. öxsxp ve (ö:p)xs desenleri karşılaştırıldığında, (ö:p)xs deseni ile kestirilen bağıl ve mutlak hata varyanslarının öxsxp desenine göre daha küçük olduğu, dolayısıyla G ve Phi katsayılarının daha büyük değerler aldığı görülmüştür. Bu iki desende yapılan Karar çalışmaları incelendiğinde, her iki desende de puanlayıcı sayısını artırmanın G ve Phi katsayılarında artış sağladığı görülmüştür. Kitapçıklarda madde sayısı sabit tutularak, puanlayıcı sayısının artırılmasının zaman, işgücü ve ekonomi açısından uygun olacağı sonucuna ulaşılmıştır.

Yıldıztekin (2014), çalışmasında, 84 öğrencinin matematik problem çözme becerisini ölçmeye yönelik hazırlanan altı adet açık uçlu sorudan oluşan bir teste verdikleri cevapları, beş puanlayıcı tarafından analitik ve bütünsel dereceli iki ayrı puanlama anahtarı kullanılarak 20-25 gün arayla puanlanması sağlanmıştır. Elde edilen verilerin, KTK kapsamında Pearson momentler çarpım korelasyon katsayısı, Spearman sıra farkları korelasyon katsayısı, Cronbach Alfa, Kappa ve Krippendorf Alfa katsayıları ile güvenilirlik analizleri yapılmıştır. Genellenebilirlik kuramında ise bxmxxp deseni çalışılmıştır. Araştırma sonuçlarına göre, KTK ve G Kuramından elde edilen güvenilirlik katsayılarının birbirleriyle tutarlı ve oldukça yüksek oldukları görülmüştür. Genellenebilirlik Kuramının bxmxxp deseninden elde edilen sonuçlara göre, puanlayıcıların analitik ve bütünsel puanlama anahtarlarıyla verdikleri puanların değişkenliğe etki etmedikleri belirlenmiştir. Ayrıca, puanlayıcılar arası tutarlılığın yüksek olduğu belirlenmiştir.

Nalbantoğlu, Yılmaz ve Başusta (2015), çalışmalarında Tıp eğitimindeki öğrencilerin mesleki becerilerinden dikiş atma ve alma becerilerine ait performans puanlarının güvenilirliğini belirlemek ve puanlama güvenilirliğini öğrenci, puanlayıcı, beceri ve görev etkileşimlerini dengelenmemiş verilerde Genellenebilirlik kuramıyla incelemişler. Araştırma kapsamında, öğrencilerin puanlayıcılarla, görevlerin ise becerilerle yuvalandığı ve bu becerilerin tüm öğrenciler için ortak olduğu dengelenmemiş (ö:p)x(g:b) deseni (ö: öğrenci, p:puanlayıcı, g: görev ve b:beceri olmak üzere) kullanılmıştır. Araştırmanın sonucunda, her iki beceri bakımından puanlayıcı etkisinden kaynaklı farklılıkların olmadığı, beceri ve ilgili beceriye ait görevlerin puanlayıcıdan puanlayıcıya farklılık göstermediği tespit edilmiştir. Öğrencilerin dikiş atma ve alma becerilerine ait performanslarının belirlenmesi sürecine yönelik genellenebilirlik kuramı ile elde edilen güvenilirlik katsayıları ise kabul edilebilir düzeyde bulunmuştur.

Bağcı (2015), çalışmasında 'Matematiksel Muhakemeyi Değerlendirme Ölçeği'nden elde edilen ölçümlerin güvenilirliğinin incelenmesini amaçlamıştır. Bu amaçla ölçeğin, üç bağımsız puanlayıcı tarafından puanlanmasıyla elde edilen ölçümlerin güvenilirliği; KTK ile

Genellenebilirlik (G) kuramının çaprazlanmış ve yuvalanmış desenlerinde karşılaştırılmıştır. Genellenebilirlik kuramı için iki farklı senaryo kullanmak üzere iki desen tasarlanmıştır. Bu desenlerden birincisi, öğrenci (ö), soru (s) ve puanlayıcı (p) değişkenleri olmak üzere, öğrencilerin aynı sorular üzerinden puanlayıcıların her biri tarafından puanlandığı öxsp çapraz desendir. İkinci desen ise, her bir puanlayıcının soruların sadece bir kısmını puanlamasıyla oluşan, puanlayıcı ve soru değişkenlerinin yuvalanmış, öğrencilerin ise bu değişkenlerle çaprazlanmış olduğu öx(s:p) desendir. Araştırma sonucunda her iki kuramdan kestirilen güvenilirlik katsayıları da kabul edilebilir düzeyde bulunmuştur. öx(s:p) deseninde G çalışması sonucu kestirilen G ve Phi katsayıları öxsp deseninden daha yüksek bulunmuştur. KTK' de her 3 puanlayıcı için ayrı ayrı hesaplanarak elde edilen Cronbach alfa katsayıları ise, her iki desende bağıl ölçme için kestirilen G katsayıları ile oldukça paralel bulunmuştur. Ayrıca Genellenebilirlik kuramında yapılan karar çalışması ile de yüzey sayılarının mutlak ve bağıl hata varyanslarına etkisi belirlenmiştir. Dolayısı ile G kuramı ile yapılan analizlerin KTK' ye göre daha detaylı bilgi verdiği görülmüştür.

Pekin (2015), "Otizm Sosyal Beceriler Profili" (OSBP) Ölçeğinin beş puanlayıcı tarafından puanlanması ile elde edilen veriler ile KTK ve Genellenebilirlik Kuramı'nda puanlayıcılar arası güvenilirlik karşılaştırmıştır. Genellenebilirlik Kuramı'nda puanlayıcıların birlikte ve dönüşümlü puanlama yapmasıyla oluşturulan farklı desenlerden ve KTK' den elde edilen güvenilirlik katsayılarının düzeyleri saptanmış ve hangi kuramın daha fazla bilgi sunduğu belirlenmeye çalışılmıştır. Elde edilen veriler KTK' de puanlayıcıların iç tutarlılık güvenilirliği Cronbach Alfa (α) katsayısı, puanlayıcılar arası güvenilirlik Kendall'ın uyum katsayısı, puanlayıcılar arası korelasyon katsayısı ve puanlayıcıların verdikleri puanlar arasında fark olup olmadığı ise ilişkili örneklemelerde varyans analizi ile hesaplanmıştır. Genellenebilirlik kuramı kapsamında ise puanlayıcıların bireyleri birlikte ve dönüşümlü puanladıkları iki farklı senaryo üzerinden iki desen tasarlanmıştır. Bireylerin (b) aynı maddeler (m) doğrultusunda puanlayıcıların (p) her biri tarafından puanlandığı bmxp çapraz deseni ve bireylerin tüm maddeler doğrultusunda farklı puanlayıcılar tarafından puanlandığı (p:b)xm yuvalanmış deseni için ayrı ayrı G ve K çalışmaları yapılmış ve sonuçlar birbirleriyle karşılaştırılmıştır. Araştırma sonucunda KTK' ye göre Cronbach Alfa katsayılarının yüksek değerler aldığı, ancak puanlayıcılar arası korelasyon katsayılarının ve Kendall'ın uyum katsayısının düşük değerler aldığı görülmüştür. Genellenebilirlik kuramı kapsamında çapraz ve yuvalanmış desen için ayrı ayrı hesaplanan G ve Phi katsayılarının kabul edilebilir düzeylerde değerler aldığı, birey ve puanlayıcı ortak etkilerine ait varyans bileşenlerinin her iki desende de toplam varyansı açıklama oranlarının yüksek olduğu görülmüştür. Her iki desende yapılan karar çalışmaları sonucunda güvenilirliği artırmak için puanlayıcı sayısını artırmanın, madde sayısını artırmaya göre daha etkili olduğu sonucuna varılmıştır.

Şalgam (2016)'ın çalışmasında ilköğretim 7.sınıf öğrencilerine yönelik olarak hazırlanmış 'Kısa Cevaplı Matematik Yazılı Sınavından' elde edilen ölçümlerin güvenilirliğinin incelenmesi amaçlanmıştır. Yazılı sınavın üç puanlayıcı tarafından puanlanmasıyla elde edilen ölçümler KTK' nin test tekrar test yöntemi ve Genellenebilirlik Kuramının çaprazlanmış deseniyle kıyaslanmıştır. Kısa Cevaplı Matematik Yazılı Sınavı iki hafta arayla aynı öğrencilere uygulanmış ve aynı puanlayıcı tarafından puanlanmıştır. Araştırma sonucunda her iki kuramdan elde edilen güvenilirlik katsayıları kıyaslandığında aralarında manidar bir farklılık olmadığı gözlenmiştir.

Aktaş, Başusta ve Alıcı (2016), çalışmalarında farklı puanlama anahtarlarının güvenilirliğini Genellenebilirlik Kuramı ile incelemişlerdir. Puanlama anahtarlarının sabit yüzey olarak belirlendiği tümüyle çaprazlanmış hpxpg (h:hikaye, p:puanlayıcı, g:görev) deseni ile G çalışması yürütülmüştür. Varyans bileşenlerinin kestirilmesinin ardından, görev ve puanlayıcı yüzey koşulları değişimlenerek Karar Çalışması gerçekleştirilmiştir. Sonuç olarak her iki puanlama anahtarında da, ana etkiler içerisinde hikayelere ilişkin değişkenlik kaynağının, ortak etkiler içerisinde ise "hp" ortak etkisinin yüksek açıklanan varyans yüzdesine sahip olduğu gözlenmiştir. Desen çözümlemesinde açıklanan en büyük varyans "hpg" ortak etkisine ait olduğu belirtilmiştir. Bu durum, ortak etkileşimin yanı sıra desende yer almayan çeşitli değişkenlik ve hata kaynaklarının varlığının göstergesi olabileceği ifade edilmiştir. Her iki puanlama anahtarı için hesaplanan G ve Phi katsayıları incelendiğinde, analitik rubrik ile puanlamanın daha güvenilir ve genellenebilir olduğu belirlenmiştir. Kontrol listesi için G ve Phi katsayılarının ikisi de 0,62; analitik rubrik için ise 0,72 olarak bulunmuştur.

1.1.11.2. Yurt Dışında Yapılan Çalışmalar

Bell (1985), çalışmasında genellenebilirlik kuramında çok büyük verilerle, dengelenmemiş (unbalanced) ve karışık desen tasarımlarında kullanılan programların sınırlılıklarına çözüm aramıştır. Çalışmada öğrenci (p), okul (s) ve soru (q) ile gösterilmek üzere farklı okullardan gelen öğrencilerin kullanıldığı (p:s)xq deseni ile dengelenmiş (balanced) ve dengelenmemiş (unbalanceed) desenler için analizler yapmıştır. Çalışmada MIVQUE metodunun bazı sınırlılıkları ortadan kaldırdığı ve dengelenmiş desenler için SAS veya GENOVA programlarını kullanmanın daha uygun olduğu belirtilmiştir.

Lane ve Sabers (1989), öğrencilerin yazma becerilerini değerlendirmede puanlama sisteminin güvenilirliğini belirlemek amacıyla genellenebilirlik kuramını kullanmışlardır. Araştırmada 15 öğrenciden verilen bir konu hakkında kompozisyon hazırlamaları istenmiştir. Öğrencilerin kağıtları 8 öğretmen tarafından bir puanlama sistemi oluşturularak değerlendirilmiştir. Öğretmenler puanlama sistemini dört kategoride hazırlamıştır. Araştırmada

öğrenciler (s), puanlayıcılar (r) ve puanlama sistemindeki kategoriler (c) ile gösterilmek üzere kategori değişkeni sabit olarak belirlenip tüm değişkenlerin çapraz tasarlandığı $s \times r \times c$ deseniyle G ve K çalışmaları yapılmıştır. Araştırma sonuçlarına göre, puanlayıcı sayısı azaldıkça mutlak ve bağıl hata varyanslarının arttığı, G ve Phi katsayılarının ise azaldığı tespit edilmiştir.

Chang ve Hocevar (2000), çalışmalarında fakültede bulunan öğrencileri ve dersi değerlendirdikleri üniversite değerlendirme verilerini Genellenebilirlik Kuramı ile farklı desenler üzerinde kullanılarak 3 G çalışması ve 3 Karar çalışması yapmışlardır. Çalışmada değişkenlik kaynağı olarak öğrenci (s), öğretmen (t), ders (c), madde (i), eğitim durumu (e), dönem (o) gibi değişkenleri ele almışlardır. Bu değişkenlik kaynaklarına göre de $(s:c):(t:ox)$, $(s:t):(c:ox)$ ve $s:(e:ox)$ desenlerini kullanarak G ve K çalışmaları yapmışlardır. Çalışma sonucunda, değerlendirme için öğrenci sayısının 10 ile 20 arasında olmasının yeterli olduğu tespit edilmiştir.

Ferguson ve Kreiter (2001), tıp fakültesi öğrencilerinin klinik değerlendirmelerinde staj gruplarının (nöroloji, pediatri, dahiliye, cerrahi ve jinekoloji) ve bu stajlardaki gözlemlerin güvenilirliğini araştırmışlardır. Çalışmada öğrenci (p), puanlayıcı (r) ve staj grupları (c) ile gösterilmek üzere araştırmacılar genellenebilirlik kuramını kullanarak öğrencilerin staj gruplarıyla çapraz, her bir stajda da farklı puanlayıcının olduğu $r:(p \times c)$ deseni ve her bir stajda puanlayıcıların etkisini araştırmak içinde $r:p$ deseni üzerinde G ve K çalışmaları yapmışlardır. Araştırma sonucunda öğrencilerin performansının stajlar değiştikçe farklılaştığı, bir stajda gözlem sayısını artırmanın G katsayısını da artırdığı tespit edilmiştir.

James, Kreiter ve diğerleri (2001), çalışmalarında genellenebilirlik kuramını kullanarak tıp öğrencilerinin kullandığı klinik eğitim sitelerinin güvenilirliğini belirlemek amaçlı G ve K çalışmaları yapmışlardır. Çalışmada kullanılan siteler 3 yapıyla değerlendirilmiştir. Bunlar sitelerin yeterlilikleri, sitelerdeki bilgilere ulaşılabilirlik ve öğrencilerin yeterlilikleridir. Çalışmada öğrencilere bu üç bölüm için ayrı değerlendirme ölçekleri verilmiştir ve her bir bölüm ayrı ayrı genellenebilirlik kuramı ile analiz edilip G ve K çalışması yapılmıştır.

Ludlow (2001), öğrencilerin okuma yetenekleriyle ilgili yaptığı çalışmasında değişken kaynağı olarak öğrenci (s), puanlayıcı (r) ve pasaj (p) değişkenlerini alarak değişkenlerin bütünüyle çaprazlanmış $s \times r \times p$ ve her bir puanlayıcının farklı pasajları değerlendirdiği $s \times (p:r)$ desenlerini kullanarak Genellenebilirlik çalışması ve iki Karar çalışması yapmıştır. Çalışma sonuçlarından elde edilen bulgulara göre güvenilir sonuçlar elde edebilmek için 2 puanlayıcı tarafından puanlanan her öğrencinin minimum üç pasaj okuması gerektiği tespit edilmiştir.

Gorter ve arkadaşları (2002), gerçek hasta kullanarak klinik yeterlilikte önem taşıyan standart hasta sınavının güvenilirliğini Genellenebilirlik Kuramı ile incelemiştir. Araştırmada 22 romatolog ve her romatologun incelediği 8 farklı hastalık öyküsü bulunmaktadır. Her hastalık öyküsünden 2 adet hasta bulunmaktadır. Romatologlar bu iki hastadan birini muayene

etmektedir. Romatologların performansı önceden hazırlanmış bir kontrol listesiyle tek bir romatolog tarafından her vaka için ayrı ayrı değerlendirilmiştir. Çalışmada Genellenebilirlik kuramıyla romatologların ve hastalık öykülerinin çaprazlandığı desen kullanılmıştır. Araştırma sonuçlarına göre, hastalık öykülerinin zor olduğunu tespit edilmiştir. Ayrıca bu çalışmayla gerçek hasta kullanılarak yapılan bu sınavın diğer sınavlarla yakın sonuç verdiği ileri sürülmüştür.

Sharma ve Weathers (2003), önceden kullanılan bir ölçeğin ülkeler arasında genellenebilirliğini tespit etmeye ve bu ölçekle ilgili K çalışması yaparak da ölçekteki madde sayısını kestirmeye çalışmışlardır. Ayrıca araştırmacılar yaptıkları bu çalışmayla ülkeler için farklı sayıda denekler kullanarak Genellenebilirlik kuramında dengelenmemiş (unbalanced) tasarımı örneklendirmeye çalışmışlardır. Bunların yanında araştırmada faktör analizini de kullanarak ölçekteki madde sayısını belirlemiş ve bu sonuçları G kuramı sonuçları ile karşılaştırmışlardır. Araştırma sonuçlarına göre, faktör analizinin ölçek içeriği hakkında ve ölçek geliştirme çalışmalarında G kuramına göre daha açıklayıcı bilgiler verdiği, G kuramı ve faktör analizinin ikisinin de ölçeğin genellenebilirliği hakkında bilgi verdiği, bunların yanında genellenebilirlik kuramının da karar çalışması yapılarak gelecek çalışmalar için ön tahminde bulunulabilmesi açısından daha önemli olduğu tespit edilmiştir. Ayrıca araştırmacılar iki kuramında sınırlılıklarının yanında faydalarının bulunduğunu, iki kuramın da birbirini tamamlayıcı olduğunu belirtmişlerdir.

Wang (2005), çalışmasında geleneksel yöntem ve G kuramıyla güvenilirlik hesaplayarak bir karşılaştırma yapmıştır. Bu amaçla öğrencilerin İngilizce yazma becerilerini ölçmek için hazırlanmış sınav verileri kullanılmıştır. Çalışmada öğrenci (p), madde (i) ve beceri (t) olmak üzere $p \times (i:t)$ deseniyle G çalışması yapılmıştır. Ayrıca araştırmacı çalışmada α güvenilirlik katsayısını hesaplamış ve bu değeri G katsayısıyla karşılaştırmıştır. Çalışmanın sonucunda α ve G katsayısının birbirine yakın sonuç verdiği yorumuna varılmıştır.

Thiessen (2006), bir matematik yerleştirme sınavında öğrenciler hakkında doğru yerleştirme kararları verebilmek için testte ihtiyaç duyulan soru sayısını belirlemek amaçlı yaptığı çalışmasında G kuramını kullanmıştır. Çalışmaya katılan öğrencilerden 50'si fen, 51'i sosyal bilimler ve 57'si de alanlarına karar veremeyen öğrencilerden oluşmaktadır. Sınav cebir, geometri ve fonksiyon içerik kategorilerinden oluşan 30 sorudan oluşmaktadır. Araştırmacı desenin dengelenmemiş (unbalanced) desen olmaması için farklı alanlardaki öğrenci sayılarını 50 olarak belirlemiştir. Araştırma sonuçlarına göre yapılan K çalışmasında yerleştirme testinin üç alanında da (cebir, geometri ve fonksiyonlar) en az 8 maddenin olması gerektiği belirlenmiştir.

Iramaneerat ve Yudkowsky (2007), çalışmalarında dört tip puanlayıcı hatasını araştırmak için 173 dördüncü sınıf tıp öğrencisinin klinik beceri değerlendirmesinde elde edilen

sınav verilerini kullanarak G kuramı ve Rasch modelini kıyaslamış ve sınav için bu iki modelin tamamlayıcı etkilerini de kullanarak sınavı geliştirmeye çalışmışlardır. Çalışmada G kuramı çeşitli kaynaklardan gelen hata varyanslarını sağlarken, Rasch analizi istasyonların tutarlılığı hakkında bilgi sağlamıştır.

Lee ve Kantor (2007), TOEFL sınavı yazma becerilerinin ölçüldüğü bölüm için alternatif değerlendirme yöntemleri ve görev örnekleri belirlemek amaçlı Genellenebilirlik kuramını kullanmışlardır. Çalışmada farklı yazma görevleri ve puanlayıcıların yazma becerilerine ait puanlardaki güvenilirliği araştırarak farklı senaryolara bağlı Karar çalışması ile, maksimum puan güvenilirliği için puanlayıcı ve görev sayısını belirlemeye çalışmışlardır. Araştırma sonuçlarına göre, maksimum puan güvenilirliği için sınavda kullanılan yazma görevlerinin sayısını artırmanın görev başına düşen puanlayıcı sayısını artırmaktan daha etkili olduğu tespit edilmiştir. Ayrıca çalışmada dinleme-yazma görevleri sayısının, okuma-yazma görevleri sayısından daha çok tutulduğunda daha yüksek puan güvenilirliği elde edilmiştir.

Au, Prahardhi ve Shiell (2008), çalışmalarında ekonomik değerlendirmeler için hazırlanmış iki ölçeğin güvenilirliğini araştırmışlardır. Ölçeklerden birincisi 16 maddelik likert tipi bir ölçektir. Diğeri ise 57 maddelik bir kontrol listesidir. Çalışmada 1990 ve 2003 yılları arasında yapılan 30 makale kullanmışlardır. Bu makaleler her iki ölçekle 2 yazar tarafından puanlanmıştır. 2 yazarın puanlayıcı güvenilirliği sınıf içi korelasyon katsayısıyla belirlenmiştir. Araştırmada makale (a), puanlayıcı (r) ve ölçek (i) olarak gösterilmek üzere G kuramıyla değişkenlerin çapraz tasarlandığı axrxi deseni analiz edilmiştir. Araştırma sonuçlarına göre yazarlar arasındaki uyum yüksek ve iki ölçeğin de eşit derecede iyi olduğu belirlenmiştir. İlk ölçeğin daha kısa olduğu, böylelikle uygulamada daha az zaman aldığı, bununla birlikte ikinci ölçekten daha az bilgi sağlandığı belirlenmiştir.

Christ ve diğerleri (2010), altı okul öncesi çocuğunun çözölemeyen bir lego ile uğraşmalarını içeren üçer dakikalık videoları, davranışları doğrudan gözlemleyen 125 puanlayıcı tarafından puanlanarak çaprazlanmış desenlerle Genellenebilirlik çalışmaları yapmışlardır. 6'lı, 10'lu ve 14'lü metrik ölçeklerle yapılan çalışmada 10'lu metrik ölçekten elde edilen puanların daha etkili ve duyarlı sonuçlar verdiğini bulmuşlardır.

1.2. Amaç

Bu araştırmanın amacı, G Kuramı'nda performansa dayalı durum belirlemede elde edilen ölçümler için farklı ve çok hata kaynaklı bir ölçme durumu olarak çok yüzeyli farklı desen kurgulamaları ile elde edilen puanların G ve Phi katsayılarının incelenmesidir.

Dört değişkenli dengelenmiş veri yapılarında G kuramının nasıl kullanılabilceğini göstermek, araştırma kapsamında kullanılan geometri puanlama anahtarının güvenilirliğini çeşitli çok yüzeyli farklı desen kurgulamalarında incelemek ve birey, okul, puanlayıcı ve görev değişkenlik kaynaklarını ve etkileşimlerini dikkate alarak üç yüzeyli desen uygulamasında Genellenabilirlik katsayıları elde etmek ve yorumlamak amaçlanmaktadır.

1.3. Problem Cümlesi

G kuramına göre; farklı okullarda bulunan öğrencilerin geometri performansının puanlanmasında kullanılan görevlerin, birden fazla puanlayıcı tarafından puanlanmaları ile oluşturulan farklı desenlerin Genellenebilirlik (G) ve Phi katsayıları sonuçlarının değişimi nasıldır?

1.3.1. Alt Problemler

1) Bireyin (b) okul (o) yüzeyinde yuvalandığı, görev (g) ve puanlayıcı (p) değişkenlerinin ise çapraz tasarlandığı (b:o) x g x p deseninin G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?

2) Bireyin (b) okul (o) yüzeyinde ve görevin (g) puanlayıcı (p) yüzeyinde yuvalandığı ve bu ikisinin çapraz tasarlandığı (b:o) x (g:p) deseninin G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?

3) Puanlayıcının (p), okul (o) yüzeyinde yuvalandığı, birey yüzeyinin de puanlayıcı (p) yüzeyinde yuvalandığı ve görev (g) yüzeyinin çapraz tasarlandığı (b:(p:o)) x g deseninin G çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıldır?

1.4. Önem

Eğitim ve psikolojideki alanyazın incelendiğinde, G Kuramı ile ilgili olan çalışmalarda, G kuramının KTK ile karşılaştırılmasına yönelik çalışmaların ağırlıkta olduğu ve çalışmalarda değişkenlik kaynaklarının çapraz tasarlandığı desenlerin sıklıkla kullanıldığı görülmektedir. G kuramındaki çok yüzeyli çalışmalara yönelik yapılan taramalarda ise; çok yüzeyli çalışmaların çoğunluğunda iki yüzeyli desenler (genellikle maddelerin ve puanlayıcıların yüzey olarak tercih edildiği) ele alınmıştır (Atılğan ve Tezbaşaran 2005, Çakıcı ;2011, Büyükkıdık , 2012; Alkan, 2013). Ülkemizde çok yüzeyli desenler kapsamında ikiden fazla yüzeyin bulunduğu çalışmalar oldukça sınırlıdır (Anıl ve Büyükkıdık, 2012; Büyükkıdık, 2012; Nalbantoğlu ve Başusta, 2015;

Aktaş, Başusta ve Alıcı, 2016). Aynı zamanda ikiden fazla yüzeyin bulunduğu ve farklı kurgulanmış desenlerin karşılaştırıldığı herhangi bir çalışmaya ise rastlanmamıştır. Genellenebilirlik çalışmalarında ikiden fazla yüzeyin bulunduğu az sayıda tez çalışması da bulunmaktadır (Büyükkıdık, 2012). Bu çalışmada üç yüzeyli desenlerin incelenmesinin bu anlamda alanyazında önemli bir boşluğu doldurması beklenmektedir. Yurt dışında yapılan G Kuramıyla ilgili çalışmalar incelendiğinde ise, değişkenlerin çapraz tasarlandığı desenlerle G ve K çalışmalarının yapılmasının ağırlık kazandığı, G Kuramının performans değerlendirme çalışmalarında güvenilirlik belirlerken ve ölçeklerdeki madde sayısının belirlenmesinde sıklıkla kullanıldığı görülmektedir.

Bu bağlamda; bu çalışmada çok yüzeyli farklı desenlerin (çapraz ve yuvalanmış) kullanılacak olması, ayrıca ölçme objesi (birey), görev ve puanlayıcı yüzeylerinin yanı sıra farklı ve önemli bir değişkenlik kaynağı olduğu düşünülen "okul" yüzeyine de yer verilmesi sebebiyle araştırma sonuçlarının alanyazına önemli bir katkı sağlayacağı düşünülmektedir.

1.5. Sayıtlar

1) Her puanlayıcının aynı performans görevleri için bütünsel puanlama anahtarı yardımıyla yaptığı puanlamalar birbirinden bağımsızdır.

2) Puanlayıcılar her öğrencinin performans görevini aynı titizlikle ve dikkatle puanlamışlardır.

1.6. Sınırlılıklar

1.Araştırma, sadece kullanılan desenlerle sınırlıdır.

2.Araştırma sadece kullanılan performans görevi sayısı, puanlayıcı sayısı ile sınırlıdır.

3.Araştırmanın çalışma grubu Mersin ili merkez lise okullarında üniversiteye hazırlanan öğrenciler ile sınırlıdır.

BÖLÜM II

2. YÖNTEM

2.1. Araştırmanın Modeli

Bu araştırma, Genellenebilirlik Kuramı'nda performansa dayalı durum belirlemede elde edilen ölçümler için çok yüzeyli farklı desen kurgulamaları ile elde edilen puanların Genellenebilirlik (G) ve Phi katsayıları incelendiğinden, betimsel bir araştırma niteliğini taşımaktadır. Betimsel araştırmalar, olanı olduğu gibi saptamaya/tanımlamaya çalışan, farklılık ya da ilişki bulmaya çalışmayan, araştırma türüdür (Erkuş, 2011).

2.2. Çalışma Grubu

Araştırmanın çalışma grubunu 2014-2015 öğretim yılında Mersin ili merkez liselerinde okuyan 128 12. sınıf öğrencisi oluşturmaktadır. Öğrenciler gönüllülük esasına dayalı olarak seçilmiştir. Araştırma sorularına yanıt aramak amacıyla lise matematik öğretmeni olarak çalışan ve araştırmaya katılmaya gönüllü 8 matematik öğretmenin katılımıyla puanlayıcı grubu oluşturulmuştur.

2.3. Veri Toplama Araç ve/veya Teknikleri

Araştırma verilerinin elde edilmesine ilişkin süreç aşağıdaki gibi gerçekleşmiştir:

Araştırma sorularının cevaplanabilmesi için iki araç geliştirilmiştir. Bunlardan ilki öğrencilere uygulanan ve sekiz geometri sorusundan oluşan açık uçlu sınavdır. İkincisi ise puanlayıcıların puanlama yapabilmesi için hazırlanan bütünsel dereceli puanlama anahtarıdır.

İlk aşamada, öncelikle 2014-2015 öğretim yılında uygulanmış bir YGS denemesinde yer alan 8 geometri sorusu seçilmiş ve yeni bir soru kağıdı hazırlanmıştır. Ek 1' de yer alan soru kağıdı, anlam ve dil bilgisi açısından iki matematik uzmanı tarafından tekrar gözden geçirilmiştir. Sorular, 2015 yılı bahar döneminde farklı iki tür (anadolu lisesi ve fen lisesi) okulda öğrenim gören, çalışma grubunda yer alan öğrencilere uygulanmıştır. Uygulama öncesi öğrencilere gerekli açıklamalar araştırmacı tarafından yapılmış ve öğrencilerin uygulama ile ilgili soruları cevaplanmıştır.

Puanlama anahtarının hazırlandığı ikinci aşamada aşağıdaki adımlar izlenmiştir.

- 1) Öncelikle soru kağıdında yer alan sorular tarafından ölçülen kazanımlar belirlenmiştir.
- 2) Öğrencilerin performanslarda göstermeleri istenen, gözlenmek istenen özel davranışlar ve gözlenmek istenmeyen davranışlar tanımlanmıştır.

- 3) Bu davranışlar arasında öne çıkan özellikler tanımlanmıştır.
- 4) Her soruda örnek cevaptan yetersiz cevaba kadar her bir özelliği dahil eden eksiksiz açıklama yazılmıştır.
- 5) Örnek cevaptan yetersiz cevaba kadar süregelen aralıktaki diğer seviyeler ortak özelliklerle tanımlanarak puanlama anahtarı tamamlanmıştır.
- 6) Hazırlanan dereceli puanlama anahtarı için ayrıca uzman görüşü alınmıştır.
- 7) Her bir geometri performans görevi 0-4 puan üzerinden ölçülmüştür. Ek 2' de belirtilen puanlama anahtarının özeti aşağıda verildiği biçimdedir:

0 puan:	Cevap Yok
1 puan:	Yetersiz Cevap
2 puan:	Kısmen Yeterli Cevap
3 puan:	Yeterli Cevap
4 puan:	Örnek cevap

- 8) Puanlayıcılara, bütünsel dereceli puanlama anahtarının ve puanlama tablolarının nasıl kullanılacağı araştırmacı tarafından açıklanmış, tüm cevap kağıtlarını puanlamaları sağlanmıştır.

Çalışma grubundaki okullarda yapılan uygulamalar tamamlandıktan sonra 8 puanlayıcıya, araştırmacı tarafından hazırlanan bütünsel dereceli puanlama anahtarı ve puanların yazılacağı "puanlama tabloları" verilmiştir. Puanlama tablolarında, her bir öğrencinin cevap kağıdına verilen, okulunu simgeleyen bir harf ve her bir öğrenci için ayrı ayrı verilmiş bir numaradan oluşan kodlar bulunmaktadır. Diğer bir deyişle, her bir öğrencinin, isminin yer almadığı, bir harf ve bir sayıdan oluşan kod numarasının karşısına her bir madde için verdiği cevaba karşılık gelen puanların yazılacağı tablolar oluşturulmuştur. Bu puanlama tablosuna ilişkin örnek Ek 3 de yer almaktadır.

2.4. Verilerin Analizi

8 geometri performans görevinin 8 puanlayıcı tarafından bütünsel puanlama anahtarı yardımıyla puanlanmasından elde edilen veriler, G Kuramı kapsamında analiz edilmiş ve elde edilen bulgular incelenmiştir. Araştırmanın gerçekleştirilmesinde aşağıdaki adımlar izlenmiştir.

1. 8 geometri performans görevi her bir okulda 64 öğrenci olmak üzere farklı düzeydeki iki okulda toplam 128 öğrenci tarafından cevaplanmıştır.

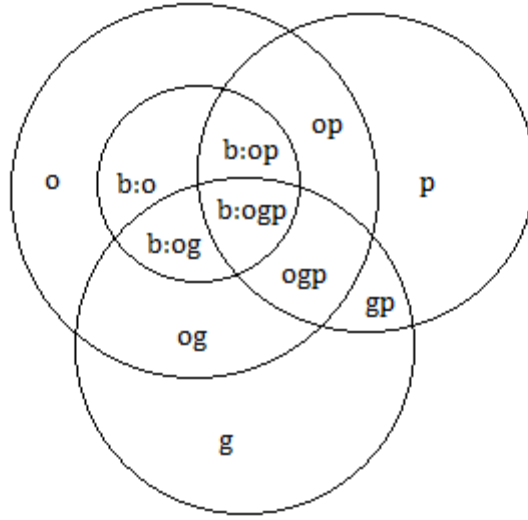
2.128 öğrencinin cevapladığı performans görevleri 8 farklı puanlayıcı tarafından puanlanmıştır. Puanlama yapılırken yukarıda açıklamaları verilen bütünsel dereceli puanlama anahtarı kullanılmıştır.

3.Elde edilen veriler desenlere uygun olacak şekilde bilgisayar ortamına aktarılmıştır. G kuramında kullanılmak üzere tamamen dengelenmiş üç desen tasarlanmıştır. G Kuramı'nda analizler her bir desen için ayrı ayrı yapılmıştır. Bilgisayar ortamına aktarmada veri giriş tablolarının nasıl düzenlendiği aşağıda detaylı olarak açıklanmıştır.

a)Araştırmada bireyin (b) okul (o) yüzeyinde yuvalandığı, görev (g) ve puanlayıcı (p) yüzeylerinin ise çapraz tasarlandığı **(b:o)gxgp** desenine ait veri giriş tablosu, varyans kaynaklarına ait venn şeması ve varyans bileşenleri aşağıda sırasıyla verilmiştir.

Tablo2.1. : (b:o)gxgp Desenine Ait Veri Giriş Tablosu

(b:o)gxgp		p1				p2				p3				p4				...	p8			
		g1	g2	...	g8	g1	g2	...	g8	g1	g2	...	g8	g1	g2	...	g8	...	g1	g2	...	g8
o1	b1	x	x	...	x	x	x	x	x	x	x	x	x	x	x	x	...	x	x	x	x	
	b2	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	...	x	x	x	x	
	
	
	
	b64	x	x	x	x	x	x	x	x	x	x	x	x	x	x	...	x	x	x	x		
o2	b65	x	x	x	x	x	x	x	x	x	x	x	x	x	x	...	x	x	x	x		
	b66	x	x	x	x	x	x	x	x	x	x	x	x	x	x	...	x	x	x	x		
		
		
		
	b128	x	x	x	x	x	x	x	x	x	x	x	x	x	...	x	x	x	x			



Şekil 2.1. (b:o)xgxp Desenine Ait Varyans Kaynakları

Şekil 2.1.'de yer alan venn şemasına göre (b:o)xgxp desenine ilişkin varyans bileşenleri ve tanımları 11 nolu eşitlikte verilmiştir:

$$\sigma^2_{(X_{bopg})} = \sigma^2_o + \sigma^2_p + \sigma^2_g + \sigma^2_{op} + \sigma^2_{bo} + \sigma^2_{gp} + \sigma^2_{og} + \sigma^2_{bop} + \sigma^2_{bog} + \sigma^2_{pog} + \sigma^2_{bopg,e} \quad (11)$$

σ^2_o : Okullarda öğrencilerin performans görevi bakımından farklılaşması

σ^2_p : Puanlayıcıların öğrenci cevaplarını puanlamada benzerlik veya farklılıkları

σ^2_g : Performans görevlerinin zorluk kolaylık bakımından farklılaşması

σ^2_{op} : Her bir okuldaki öğrencileri puanlayan puanlayıcıların performansının okuldan okula farklılığı

σ^2_{bo} : Aynı okulda bulunan bireylerin geometri becerileri bakımından okullardaki farklılaşması

σ^2_{gp} : Puanlayıcıların bireyleri bir görevden diğerine puanlamadaki farklılaşma

σ^2_{og} : Okullardaki öğrenciler için performans görevleri arasındaki farklılaşma

σ^2_{bop} : Bireylerin geometri becerileri bakımından puanlayıcılar tarafından puanlanmasının okuldan okula farklılaşması

σ^2_{bog} : Öğrencilerin geometri becerilerinin bir okuldan diğerine farklılaşması

σ^2_{pog} : Okul, görev, puanlayıcı etkileşimi

$\sigma^2_{bopg,e}$: Artık varyans

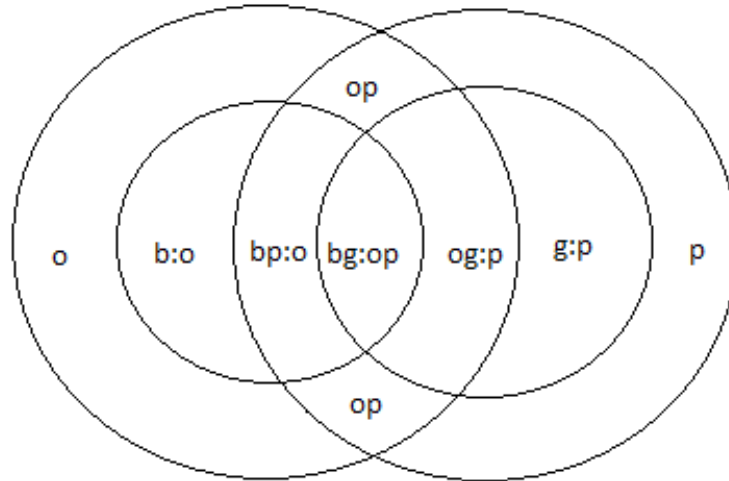
b) Bireyin (b) okul (o) yüzeyinde ve görevin (g) puanlayıcı (p) yüzeyinde yuvalandığı ve bu ikisinin çapraz tasarlandığı **(b:o)x(g:p)** desenine ait veri giriş tablosu, varyans kaynaklarına ait venn şeması ve varyans bileşenleri aşağıda sırasıyla verilmiştir.

Tablo 2.2. (b:o)x(g:p) Desenine Ait Veri Giriş Tablosu

(b:o)x(g:p)		p1		p2		p3		p4	
		g1	g2	g3	g4	g5	g6	g7	g8
o1	b1	x	x	x	x	x	x	x	x
	b2	x	x	x	x	x	x	x	x

	b64	x	x	x	x	x	x	x	x
o2	b65	x	x	x	x	x	x	x	x
	b66	x	x	x	x	x	x	x	x

	b128	x	x	x	x	x	x	x	x



Şekil 2.2. (b:o)x(g:p) Desenine Ait Varyans Kaynakları

Şekil 2.2.'de yer alan venn şemasına göre (b:o)x(g:p) desenine ilişkin varyans bileşenleri ve tanımları 12 nolu eşitlikte verilmiştir:

$$\sigma^2_{(X_{bopg})} = \sigma^2_o + \sigma^2_p + \sigma^2_{op} + \sigma^2_{bo} + \sigma^2_{gp} + \sigma^2_{bpo} + \sigma^2_{ogp} + \sigma^2_{bgop,e} \quad (12)$$

σ^2_o : Okullarda öğrencilerin performans görevi bakımından farklılaşması

σ^2_p : Puanlayıcıların öğrenci cevaplarını puanlamada benzerlik veya farklılıkları

σ^2_{op} : Her bir okuldaki öğrencileri puanlayan puanlayıcıların performansının okuldan okula farklılığı

σ^2_{bo} : Aynı okulda bulunan bireylerin geometri becerileri bakımından okullardaki farklılaşması

σ^2_{gp} : Aynı puanlayıcılar tarafından puanlanan performans görevlerindeki farklılaşma

σ^2_{bpo} : Aynı okulda bulunan bireylerin geometri becerileri bakımından puanlayıcılar tarafından puanlanmasının okuldan okula farklılaşması

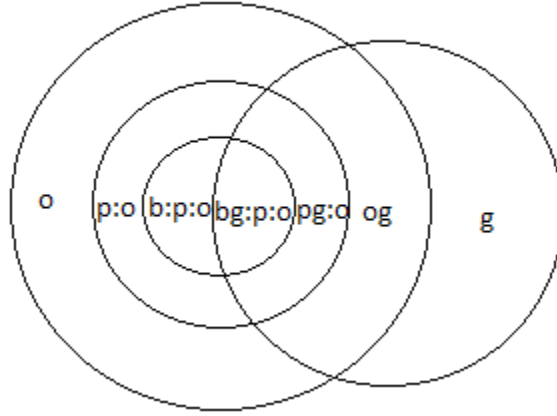
σ^2_{ogp} : Okul, görev, puanlayıcı etkileşimi

$\sigma^2_{bgop,e}$: Artık varyans

c) Puanlayıcının (p), okul (o) yüzeyinde yuvalandığı, birey yüzeyinin de puanlayıcı (p) yüzeyinde yuvalandığı ve görev (g) yüzeyinin çapraz tasarlandığı **(b:(p:o))xg** desenine ait veri giriş tablosu, varyans kaynaklarına ait venn şeması ve varyans bileşenleri aşağıda sırasıyla verilmiştir.

Tablo 2.3. (b:(p:o))xg Desenine Ait Veri Giriş Tablosu

(b: (p:o))xg		g1	g2	g3	g4	g5	g6	g7	g8	
o1	p1	b1	x	x	x	x	x	x	x	x
		b2	x	x	x	x	x	x	x	x
	
		b16	x	x	x	x	x	x	x	x
	p2	b17	x	x	x	x	x	x	x	x
		b18	x	x	x	x	x	x	x	x
	
		b32	x	x	x	x	x	x	x	x
	p3	b33	x	x	x	x	x	x	x	x
		b34	x	x	x	x	x	x	x	x
	
		b48	x	x	x	x	x	x	x	x
	p4	b49	x	x	x	x	x	x	x	x
		b50	x	x	x	x	x	x	x	x
	
		b64	x	x	x	x	x	x	x	x
o2	p5	b65	x	x	x	x	x	x	x	x
		b66	x	x	x	x	x	x	x	x
	
		b80	x	x	x	x	x	x	x	x
	p6	b81	x	x	x	x	x	x	x	x
		b82	x	x	x	x	x	x	x	x
	
		b96	x	x	x	x	x	x	x	x
	p7	b97	x	x	x	x	x	x	x	x
		b98	x	x	x	x	x	x	x	x
	
		b112	x	x	x	x	x	x	x	x
	p8	b113	x	x	x	x	x	x	x	x
		b114	x	x	x	x	x	x	x	x
	
		b128	x	x	x	x	x	x	x	x



Şekil 2.3. (b:(p:o))xg Desenine Ait Varyans Kaynakları

Şekil 2.3.'de yer alan venn şemasına göre (b:(p:o))xg desenine ilişkin varyans bileşenleri ve tanımları 13 nolu eşitlikte verilmiştir:

$$\sigma^2_{(X_{pgbo})} = \sigma^2_o + \sigma^2_g + \sigma^2_{po} + \sigma^2_{og} + \sigma^2_{bpo} + \sigma^2_{pgo} + \sigma^2_{bgpo,e} \quad (13)$$

σ^2_o : Okullarda öğrencilerin performans görevi bakımından farklılaşması

σ^2_g : Performans görevlerinin zorluk kolaylık bakımından farklılaşması

σ^2_{po} : Aynı okuldaki öğrencileri puanlayan puanlayıcıların performansının okuldan okula farklılığı

σ^2_{og} : Farklı okullarda bulunan bireylerin geometri becerileri bakımından okullardaki farklılaşması

σ^2_{gp} : Farklı puanlayıcılar tarafından puanlanan performans görevlerindeki farklılaşma

σ^2_{bpo} : Aynı okulda bulunan bireylerin geometri becerileri bakımından aynı puanlayıcılar tarafından puanlanmasının okuldan okula farklılaşması

σ^2_{pgo} : Okul, görev, puanlayıcı etkileşimi

$\sigma^2_{bgpo,e}$: Artık varyans

4. Bu üç farklı desene bağlı olarak yapılan puanlamaların güvenilirliği G ve Phi katsayıları hesaplanarak incelenmiştir.

G kuramı ile desenlere ait varyans bileşenlerinin kestirilmesi, değişkenlerin toplam varyansı açıklama oranlarının hesaplanmasında EduG 6 programı kullanılmıştır. EduG 6 programı, G kuramı analizleri için geliştirilmiş olup, araştırmacının oluşturduğu desenler için tanımlamış olduğu değişkenlik kaynaklarıyla G ve K çalışması yapılmasına olanak sağlamaktadır (Güler, Kaya Uyanık ve Taşdelen Teker, 2012).

5. Genellenabilirlik çalışması sonucunda kestirilen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri, G ve Phi katsayıları herbir desen için ayrı ayrı verilmiştir. Analiz çıktıları Ek 4'te yer almaktadır.



BÖLÜM III**3. BULGULAR**

Bu bölümde, araştırma sorularının her biri için yapılan analizler sonucu elde edilen bulgulara yer verilmiştir.

3.1. Bireyin (b) Okul (o) Yüzeyinde Yuvalandığı, Görev (g) ve Puanlayıcı (p) Yüzeylerinin ise Çapraz Tasarlandığı (b:o)ıxıxp Deseninin Genellenebilirlik Çalışması Sonuçları

Araştırmada ilk alt probleme yanıt aramak için yapılan Genellenebilirlik çalışması sonunda elde edilen varyans bileşenlerine ilişkin sonuçlar Tablo 3.1. de verilmiştir.

Tablo 3.1. Bireyin (b) Okul (o) Yüzeyinde Yuvalandığı, Görev (g) ve Puanlayıcı (p) Yüzeylerinin ise Çapraz Tasarlandığı (b:o)ıxıxp Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	Varyans Yüzdesi
o	730.93372	1	730.93372	0.16445	9.8
b:o	2845.31226	126	22.58184	0.27939	16.6
p	857.05945	7	122.43706	0.10976	6.5
g	2133.90125	7	304.84304	0.26182	15.5
op	46.63757	7	6.66251	0.01151	0.7
og	233.55750	7	33.36536	0.05564	3.3
bp:o	355.03735	882	0.40254	0.02409	1.4
bg:o	3976.02563	882	4.50797	0.53727	31.9
pg	193.79309	49	3.95496	0.02639	1.6
opg	28.30090	49	0.57757	0.00575	0.3
bpg:o	1295.54663	6174	0.20984	0.20984	12.4
Toplam	730.93372	6143			100%

Tablo 3.1.'de verilen, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında b:o ana etkisine ait varyans bileşeni toplam varyansın % 16.6'sını açıklamaktadır. Varyans bileşeni $\sigma^2(b:o)$, bireylere ait varyans bileşenini ($\sigma^2(b)$) ve birey x okul ortak etkileşimini ($\sigma^2(bo)$) temsil etmektedir. Varyans bileşeni $\sigma^2(b:o)$ 'nin büyük olması, farklı okullarda bulunan bireylerin geometri becerileri bakımından farklılık gösterdiği, okullarda bulunan bireylerin ölçülen özellik bakımından benzeşik olmadığı ve birey-okul ortak etkileşiminin farklılaştığı şeklinde yorumlanabilir. Okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın % 9.8 'ini açıkladığı dikkate alındığında $\sigma^2(b:o)$ 'nin yüksek çıkmasının öğrenci farklılaşmasından kaynaklı olduğu düşünülebilir. Bu durum, farklı türde (anadolu lisesi ve fen lisesi) okullarda bulunan öğrencilerin geometri hazırbuluşluklarının farklı olmasından da kaynaklanabilir.

Görev ana etkisine ait varyans bileşeni $\sigma^2(g)$, performans görevlerinin zorluk kolaylık bakımından farklılık gösterip göstermediği hakkında bilgi verir (Shavelson ve Webb, 1991). Görev değişkenine ait varyans bileşeninin toplam varyansın % 15.5 'ini açıkladığı görülmektedir. Bu durumda, her öğrencinin cevapladığı geometri görevleri arasında farklılaşmanın fazla olduğu söylenebilir. Elde edilen bu sonuç, performans görevlerinin farklı güçlüklerde olduğu ve zorluk-kolaylık düzeylerinin değiştiği şeklinde yorumlanabilir. Nitekim Ek 1 'de yer alan soru kağıdı incelendiğinde; kağıtta yer alan geometri alt alanlarının (geometri-analitik geometri) birlikte bulunması bütün görevlerin öğrenci açısından zor ya da kolay olarak algılanmasını etkiliyor olabilir.

Okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın % 9.8 ini açıkladığı görülmektedir. Okullara ait varyans bileşeninin toplam varyansı açıklama oranının % 9.8 olması diğer ana etkiler dikkate alındığında nispeten küçük olması sebebiyle, okullarda geometri becerileri bakımından farklılaşmanın az olduğunun bir göstergesidir.

Puanlayıcı varyansı, puanlayıcıların diğer puanlayıcılara göre öğrencileri değerlendirmede cömertliklerinin veya katılıklarının değişip değişmediği hakkında bilgi verir (Shavelson ve Webb, 1991). Puanlayıcı ana etkisine ait varyans bileşeninin $\sigma^2(p)$ toplam varyansın % 6.5 sini açıkladığı görülmektedir. Bu varyans bileşenine ait yüzde, tüm ana etkilerden kaynaklı varyans yüzdelerinden nispeten küçük olması sebebiyle, puanlayıcıların öğrenci cevaplarını puanlamada benzer cömertliğe/katılığa sahip oldukları, yani geometri becerilerini puanlama bakımından puanlayıcıdan puanlayıcıya farklılaşmanın fazla olmadığı biçiminde yorumlanabilir.

Varyans bileşeni $\sigma^2(bg;o)$, birey x görev ortak etkileşimini ($\sigma^2(bg)$) ve birey x görev x okul ortak etkileşimini ($\sigma^2(bgo)$) temsil etmektedir. Varyans bileşeni $\sigma^2(bg;o)$, toplam varyansın % 31.9'unu açıklamaktadır. Bu değer büyük olması, farklı okullarda bulunan öğrencilerin geometri becerileri bakımından okuldan okula farklılık gösterdiği, birey-görev-okul ortak etkileşiminin farklılaştığı, öğrenciler için geometri performans görevlerinin bir okuldan diğerine aynı zorluk/kolaylıkta olmadığı şeklinde yorumlanabilir. Ana etkiler dikkate alındığında; b:o ana etkisine ait varyans bileşeni $\sigma^2(b;o)$ ve görev ana etkisine ait varyans bileşeni $\sigma^2(g)$ nin diğer ana etkilere göre görece büyük olması, bu etkileşimin büyük olması sonucu ile örtüşmektedir. Bir diğer ifade ile bu etkileşime ilişkin varyans bileşeninin büyük çıkması bu bileşeni oluşturan ana etkilerin varyans bileşenlerinin büyük çıkması bulgusu ile örtüşmektedir.

Tabloda yer alan varyans bileşenlerinden okul-görev ortak etkisine ait kestirilen varyans bileşenin $\sigma^2(og)$ toplam varyansı açıklama yüzdesi % 3.3'tür. Bu bulguya göre, soru kağıdında yer alan geometri performans görevlerinin zorluk düzeyi bakımından, farklı okullarda bulunan öğrenciler arasında farklılık yaratmadığı söylenebilir. Yani bütün okullarda

performans görevleri benzer güçlüktedir. Ana etkiler dikkate alındığında; söz konusu bu bulgu görev ana etkisine ait varyans bileşeni $\sigma^2(g)$ nin görece büyük olması ile örtüşmemektedir. Bu durum, okul ve görev etkileşim etkisine okul ve görev değişkenlik kaynağı dışında farklı değişkenlik kaynaklarının da etkide bulunduğunu veya farklı seçkisiz hata kaynaklarının (okullarda bulunan öğrencilerin görevlerle daha önceden karşılaşmış olmaları, görevlerin zorluk veya kolaylık bakımından farklı algılanması, okullarda ölçme ortamlarının birbirinden farklı olması vb ...) etkileşime dahil olduğunu göstermektedir.

Puanlayıcı-görev ortak etkisine ait varyans bileşenin $\sigma^2(pg)$ toplam varyansı açıklama yüzdesine bakıldığında, % 1.6 'lık bu değer küçük bir değer olduğu, dolayısıyla öğrencilerin performansları puanlanırken puanların bir puanlayıcıdan diğerine az miktarda değişiklik gösterdiği, puanlayıcıların bireyleri bir görevden diğerine kararlı puanladıkları şeklinde yorumlanabilir.

Varyans bileşeni $\sigma^2(bp:o)$, birey x puanlayıcı ortak etkileşimini ($\sigma^2(bp)$) ve birey x puanlayıcı x okul ortak etkileşimini ($\sigma^2(bpo)$) temsil etmektedir (Brennan, 2001). Varyans bileşeni $\sigma^2(bp:o)$, toplam varyansın % 1.4 'ünü açıklamaktadır. Bu değer küçük olması, okulların ölçülen özellik bakımından benzeşik olduğu ve birey-puanlayıcı-okul ortak etkileşiminin farklılaşmadığı, öğrencilerin geometri becerilerinin bir okuldan diğerine değişmediği, puanlayıcıların her bir öğrencinin geometri performans görevlerini puanlarken bir okuldan diğerine tutarlı davrandığı şeklinde yorumlanabilir. Söz konusu bu bulgu okul ana etkisine ait varyans bileşenin $\sigma^2(o)$ ve puanlayıcı ana etkisine ait varyans bileşenin $\sigma^2(p)$ diğer ana etkilere oranla küçük olması bulgusu ile örtüşmektedir.

Okul ve puanlayıcı ortak etkisine ait varyans bileşenin $\sigma^2(op)$ toplam varyansı açıklama yüzdesi % 0.7 'dir. Okul ve puanlayıcı etkileşimi her bir okuldaki öğrencileri puanlayanların puanlamalarında meydana gelen değişikliğin göstergesidir. Bu durumda puanlayıcıların bireyleri puanlamasının okuldan okula değişmediğini söylemek mümkündür.

Okul, puanlayıcı ve görev ortak etkisine ait varyans bileşenin $\sigma^2(opg)$ toplam varyansı açıklama yüzdesi 0.3 'tür. Üç değişkenlik kaynağının etkileşiminden elde edilen varyans bileşenin sıfır (0) olması istenen bir durumdur. Bu oranın sıfıra yakın olması okul, görev, puanlayıcı etkileşiminde, seçkisiz hata kaynaklarının küçük olduğunun bir göstergesidir (Shavelson ve Webb, 1991). Bunun yanı sıra, bu bulgu ölçme objesi olan bireylerin temelde farklılaşması ile de ilgili olabilir. Temel değişkenliğin istendik bir durum olan bireylerden kaynaklı olduğu söylenebilir.

Birey x puanlayıcı x görev x okul $\sigma^2(bpg:o)$, ortak etkisine ait varyans bileşeni artık varyans olarak adlandırılır. Artık varyans bileşeni toplam varyansın %12.4'ünü açıklamaktadır. Tablo incelendiğinde bu varyans bileşeni büyüklük bakımından dördüncü sırada yer almaktadır. Artık varyans bileşenin %12.4 olması puanlayıcı, görev, birey ve okul ortak etkileşimi ve/veya

seçkisiz hata kaynaklarının yanı sıra başka değişkenlik kaynaklarının da deseni etkilediği şeklinde yorumlanabilir. Bu değişkenlik kaynakları cinsiyet, puanlayıcı nitelikleri, alt beceri alanları (geometri-analitik geometri) olabilir.

3.2. Bireyin (b) Okul (o) Yüzeyinde ve Görevin (g) Puanlayıcı (p) Yüzeyinde Yuvalandığı ve Bu İkisinin Çapraz Tasarlandığı (b:o)x(g:p) Deseninin Genellenebilirlik Çalışması Sonuçları

Araştırmada ikinci alt probleme yanıt aramak için yapılan Genellenebilirlik çalışması sonunda elde edilen varyans bileşenlerine ilişkin sonuçlar Tablo 3.2. 'de verilmiştir.

Tablo 3.2. Bireyin (b) Okul (o) Yüzeyinde ve Görevin (g) Puanlayıcı (p) Yüzeyinde Yuvalandığı ve Bu İkisinin Çapraz Tasarlandığı (b:o)x(g:p) Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ort.	Varyans	Varyans Yüzdesi
o	72.25000	1	72.25000	0.13494	6.9
b:o	334.50000	126	2.65476	0.22390	11.4
p	495.47656	3	165.15885	0.52391	26.6
g:p	143.21094	4	35.80273	0.23181	11.8
op	4.10156	3	1.36719	-0.03870	0.0
og:p	24.52344	4	6.13086	0.08528	4.3
bp:o	326.42188	378	0.86355	0.09520	4.8
bg:op	339.26563	504	0.67315	0.67315	34.2
Toplam	1739.75000	1023			100%

Tablo 3.2.'de verilen, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında tüm ana etkiler içerisinde puanlayıcı ana etkisine ait varyans bileşeninin $\sigma^2(p)$ toplam varyansın % 26.6 'ünü açıkladığı görülmektedir. Puanlayıcılara ait varyans bileşeninin toplam varyansı açıklama oranının % 26.6 olması, geometri becerilerini puanlama bakımından puanlayıcıdan puanlayıcıya çok fazla farklılaşma olduğunun bir göstergesidir. Bir başka ifade ile puanlayıcıların öğrenci cevaplarını puanlamada benzer cömerliğe/katılığa sahip olmadıkları söylenebilir. Bu durumda seçilen puanlayıcıların geometri görevleri bakımından bir yaşantı geçirmesinin, uzmanlık ve ilgi alanlarının bu durumu etkilediği düşünülebilir. Etkileşim etkisinde de varyans bileşenleri açıklama yüzdeleri dikkate alındığında en büyük değer $\sigma^2(gp)$ 'ye ait olması da bu durumu desteklemektedir.

Varyans bileşeni $\sigma^2(g:p)$, görevlere ait varyans bileşenini ($\sigma^2(g)$) ve görev x puanlayıcı ortak etkileşimini ($\sigma^2(gp)$) temsil etmektedir (Brennan, 2001). Ana etki g:p için kestirilen varyans bileşeni toplam varyansın %11.8'ini açıklamaktadır. Varyans bileşeni $\sigma^2(g:p)$ 'nin nispeten büyük olması, geometri görevlerinin puanlanmasının puanlayıcıdan puanlayıcıya farklılık gösterdiği, görev puanlayıcı ortak etkileşiminin farklılaştığı ve öğrencilerin geometri

becerilerinin bir puanlayıcıdan diğerine değiştiği şeklinde yorumlanabilir. Bu bulgu, puanlayıcı ana etkisine ait varyans bileşeninin $\sigma^2(p)$ büyüklüğü de göz önüne alındığında; puanlayıcıların puanlamaya konu olan özellikle ilişkili bir durumun varlığını düşündürebilir. Şöyle ki puanlamayı yapan puanlayıcının puanlama yaptığı göreve ilişkin ilgi düzeyi, temel uzmanlığı, vs. gibi göreve ilişkin geçmiş yaşantıları puanlayıcının görevden göreve puanlamasını değiştiriyor olabilir. Ayrıca farklı puanlayıcıların farklı görevleri puanlıyor olması da (g:p) bu durumu etkiliyor olabilir. Yuvalanmış desenler söz konusu olduğunda varyans bileşenleri birbirinden bağımsız olarak yorumlanamaz (Güler, Kaya Uyanık ve Taşdelen Teker,2012). Görevlerin güçlük düzeyleri arasındaki farklılıklar yorumlanırken, bütün görevlerin bütün puanlayıcılar tarafından puanlanmamasından dolayı her bir puanlayıcının puanladığı görevler üzerinden yorumlanmasının uygun olabileceğini düşündürmektedir.

Varyans bileşeni $\sigma^2(b:o)$, bireylere ait varyans bileşenini ($\sigma^2(b)$) ve birey x okul ortak etkileşimini ($\sigma^2(bo)$) temsil etmektedir (Brennan, 2001). Ana etki b:o için kestirilen varyans bileşeni toplam varyansın %11.4'ünü açıklamaktadır. Varyans bileşeni $\sigma^2(b:o)$ 'nin tabloda yer alan diğer varyans bileşenleri dikkate alındığında görece büyük olması, öğrencilerin geometri becerilerinin bir okuldan diğerine değiştiği, okullarda performans görevlerini yanıtlayan öğrencilerin ölçülen özellik bakımından benzeşik olmadığı ve birey-okul ortak etkileşiminin farklılaştığı şeklinde yorumlanabilir.

Okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın % 6.9 'unu açıkladığı görülmektedir. Okullara ait varyans bileşeninin toplam varyansı açıklama oranının % 6.9 olması, okullarda geometri becerileri bakımından farklılaşmanın az olduğunun bir göstergesidir.

Tabloda yer alan varyans bileşenlerinden bazıları negatif değer almıştır. Varyans bileşenlerinin negatif değer alması örneklem hatasından ya da modelin yanlış belirlenmesinden dolayı ortaya çıkabilir (Shavelson ve Webb, 1991). Calkins (1978) , Leone ve Nelson (1966) yaptıkları çalışmalarda değişkenlik kaynaklarının düzeyi az sayıda ise (örn. beş) varyans bileşenlerinin negatif çıkabileceğini belirtmişlerdir. Dolayısıyla bu çalışma da bu bulgunun değişkenlik kaynaklarının koşul sayılarının az olmasından kaynaklandığı düşünülmektedir. Negatif varyans kavramsal olarak mümkün olmadığından negatif değer alan varyans bileşenleri iki farklı yaklaşımla yorumlanabilir. Bu yaklaşımlardan ilki, negatif varyans bileşeni diğer varyans bileşenleri için beklenen kareler ortalamaları denkleminde yer alıyorsa, negatif değer yerine sıfır değeri kullanılarak hesaplama yapılmasıdır (Cronbach, 1972; akt: Shavelson ve Webb, 1991). Diğer yaklaşım ise bu değeri sıfır kabul etmek ancak denkleminde negatif değeri kullanmaktır (Brennan, 1983). Bu araştırmada, ilk yaklaşım dikkate alınmıştır. Brennan ve Cronbach genellenebilirlik katsayısının hesaplanmasında kestirilen negatif varyans değeri yerine sıfır kullanılmasının uygun olduğunu vurgulamaktadırlar (Shavelson ve Webb, 1991; Brennan, 2001; Atılgan, 2004). Okul ve puanlayıcı ortak etkisine ait varyans bileşeninin, $\sigma^2(op)$

negatif olması nedeniyle, toplam varyansı açıklama yüzdesi 0 'dır. Bu durumda puanlayıcı performansının okuldan okula değişmediğini söylemek mümkündür. Puanlayıcı ana etkisine ait varyans bileşeninin $\sigma^2(p)$ toplam varyansın % 26.6' sını ve okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın % 6.9 'unu açıkladığı göz önüne alındığında okul-puanlayıcı etkileşim etkisi ile örtüşmediği görülmektedir. Bunun nedeni, puanlayıcıların okullardaki bütün öğrencilerin bütün performans görevlerini puanlamamasından kaynaklanabilir.

Varyans bileşeni $\sigma^2(bp:o)$, birey x puanlayıcı ortak etkileşimini ($\sigma^2(bp)$) ve birey x puanlayıcı x okul ortak etkileşimini ($\sigma^2(bpo)$) temsil etmektedir (Brennan, 2001). Varyans bileşeni $\sigma^2(bp:o)$, toplam varyansın % 4.8 'ini açıklamaktadır. Bu değer nispeten küçük olması, bireylerin geometri becerileri bakımından puanlayıcılar tarafından puanlanmasının okuldan okula farklılık göstermediği, okulların ölçülen özellik bakımından benzeşik olduğu ve birey-puanlayıcı-okul ortak etkileşiminin farklılaşmadığı, öğrencilerin geometri becerilerinin bir okuldan diğerine değişmediği, puanlayıcıların her bir öğrencinin geometri performans görevlerini puanlarken bir okuldan diğerine tutarlı davrandığı şeklinde yorumlanabilir.

Puanlayıcı x görev x birey x okul $\sigma^2(bg:op)$, ortak etkisine ait varyans bileşeni artık varyans olarak adlandırılır. Artık varyans bileşeni toplam varyansın %34.2'sini açıklamaktadır. Artık varyans bileşeninin büyük bulunması puanlayıcı, görev, birey ve okul ortak etkileşimi ve/veya seçkisiz hata kaynaklarının büyük olabileceğinin göstergesi olabilir. Yanı sıra bu bulgu; desende yer almayan, sonucun genellenebilirliğini etkileyebilecek farklı değişkenlik kaynaklarını da akla getirmektedir. Bu değişkenlik kaynakları; yaşantılar, puanlayıcı nitelikleri, öğrencilerin alanları (sayısal-eşit ağırlık), vb. olabilir.

3.3. Puanlayıcının (p), Okul (o) Yüzeyinde Yuvalandığı, Birey Yüzeyinin de Puanlayıcı (p) Yüzeyinde Yuvalandığı ve Görev (g) Yüzeyinin Çapraz Tasarlandığı (b:(p:o))xg Deseninin Genellenebilirlik Çalışması Sonuçları

Araştırmada üçüncü alt probleme yanıt aramak için yapılan genellenebilirlik çalışması sonunda elde edilen varyans bileşenlerine ilişkin sonuçlar Tablo 3.3. 'de verilmiştir.

Tablo 3.3. Puanlayıcının (p), Okul (o) Yüzeyinde Yuvalandığı, Birey Yüzeyinin de Puanlayıcı (p) Yüzeyinde Yuvalandığı ve Görev (g) Yüzeyinin Çapraz Tasarlandığı (b:(p:o))xg Desenine Ait G Çalışması Sonucunda Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı	Sd	Kareler Ortalaması	Varyans	Varyans Yüzdesi
o	312.84766	1	312.84766	0.58821	30.6
p:o	64.68359	6	10.78060	0.05624	2.9
b:p:o	311.71875	120	2.59766	0.23609	12.3
g	250.95313	7	35.85045	0.25980	13.5
og	18.16797	7	2.59542	0.01410	0.7
pg:o	71.09766	42	1.69280	0.06149	3.2
bg:p:o	595.53125	840	0.70897	0.70897	36.8

Tablo 3.3.'de verilen, G çalışması sonucunda kestirilen varyans ve toplam varyansı açıklama yüzdelerine bakıldığında ana etkiler büyükten küçüğe doğru o, g, b:p:o ve p:o varyans bileşenleri şeklinde sıralanmaktadır. Okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın % 30.6 'sını açıkladığı görülmektedir. Okullara ait varyans bileşeninin çok yüksek olması, okullarda geometri becerileri bakımından farklılaşmanın fazla olduğunun bir göstergesidir.

Görev ana etkisine ait varyans bileşeninin $\sigma^2(g)$, toplam varyansın % 13.5 'ini açıkladığı görülmektedir. Bu durumda, her öğrencinin cevapladığı geometri görevleri arasında farklılaşmanın fazla olduğu söylenebilir. Elde edilen bu sonuç performans görevlerinin farklı güçlüklerde olduğu ve zorluk-kolaylık düzeylerinin değiştiği şeklinde yorumlanabilir.

Varyans bileşeni $\sigma^2(b:p:o)$, bireylere ait varyans bileşenini ($\sigma^2(b)$) ve birey x puanlayıcı ortak etkileşimini ($\sigma^2(bp)$) ve birey x puanlayıcı x okul ortak etkileşimini ($\sigma^2(bpo)$) temsil etmektedir (Brennan, 2001). Varyans bileşeni $\sigma^2(b:p:o)$, toplam varyansın % 12.3 'ünü açıklayarak dördüncü sırada yer almaktadır. Bu bulgu, farklı okullarda bulunan öğrencilerin geometri becerileri bakımından okuldan okula farklılaştığı, okulların ölçülen özellik bakımından benzeşik olmadığı ve birey-puanlayıcı-okul ortak etkileşiminin farklılaştığı, farklı okullarda bulunan farklı her bir öğrencinin geometri performans görevlerini puanlarken puanlayıcının bir okuldan diğerine tutarlı davranmadığı şeklinde yorumlanabilir. Aynı zamanda bu bulgu; ana etki p:o için kestirilen varyans bileşeni dikkate alındığında varyans bileşeni $\sigma^2(b:p:o)$ yüzdesindeki artışın temel kaynağının bireyler olduğunu düşündürebilir.

Varyans bileşeni $\sigma^2(p:o)$, puanlayıcı x okul ortak etkileşimini ($\sigma^2(po)$) temsil etmektedir (Brennan, 2001). Ana etki p:o için kestirilen varyans bileşeni toplam varyansın % 2.9' unu açıklamaktadır. Bu varyans bileşeninin küçük olması, puanlayıcıların geometri becerilerini puanlama bakımından okullarda farklılık göstermediği, okulların puanlayıcıların puanladığı özellik bakımından benzeşik olduğu ve puanlayıcı-okul ortak etkileşiminin farklılaşmadığı

şeklinde yorumlanabilir. Bu varyans bileşeninin küçük olması, aynı zamanda farklı okullarda farklı puanlayıcıların görevleri puanlamasından ($p:o$) kaynaklanmış olabilir.

Varyans bileşeni $\sigma^2(pg:o)$, puanlayıcı x görev ortak etkileşimini ($\sigma^2(pg)$) ve puanlayıcı x görev x okul ortak etkileşimini ($\sigma^2(pgo)$) temsil etmektedir (Brennan, 2001). Varyans bileşeni $\sigma^2(pg:o)$, toplam varyansın % 3.2' sini açıklamaktadır. Bu değer küçük olması, puanlayıcıların geometri becerilerini puanlama bakımından okuldan okula farklılık göstermediği, okulların ölçülen özellik bakımından benzeşik olduğu ve puanlayıcı-görev-okul ortak etkileşiminin farklılaşmadığı, geometri görevlerinin güçlüklerinin bir görevden diğerine çok fazla değişmediği şeklinde yorumlanabilir.

Okul ve görev ortak etkisi için kestirilen varyans bileşeninin $\sigma^2(og)$ toplam varyansı açıklama yüzdesi % 0.7' dir. Bu durum, kitapçıkta yer alan 8 geometri performans görevinin zorluk düzeyi bakımından birbirinden farklılık göstermediği, okullar arasında farklılık bulunmadığı biçiminde açıklanabilir. Bu bulgu; okul ana etkisine ait varyans bileşeninin $\sigma^2(o)$ toplam varyansın % 30.6 ve görev ana etkisine ait varyans bileşeninin $\sigma^2(g)$, toplam varyansın % 13.5 'ini açıklaması ile örtüşmemektedir. Bu durumun nedeni; okul ve görev değişkenlik kaynakları dışında başka değişkenlik kaynaklarından kaynaklanabilir. Bu değişkenlik kaynakları; okulların ölçme ortamları, bireylerin performans görevlerine yönelik geçmiş yaşantıları olabilir.

Puanlayıcı x görev x birey x okul $\sigma^2(bg:p:o)$, ortak etkisine ait varyans bileşeni artık varyans olarak adlandırılır. Artık varyans bileşeni toplam varyansın % 36.8 'ini açıklamaktadır. Artık varyans bileşeninin büyük çıkması puanlayıcı, görev, birey ve okul ortak etkileşimi ve/veya seçkisiz hata kaynaklarının büyük olabileceğinin göstergesi olabilir.

BÖLÜM IV

4. SONUÇ, TARTIŞMA VE ÖNERİLER

$(b:o) \times g \times p$, $(b:o) \times (g:p)$ ve $(b:(p:o)) \times g$ desenlerinden G çalışmasıyla elde edilen varyans ve toplam varyansı açıklama oranları incelendiğinde;

Okul ana etkisine ait varyans bileşeninin toplam varyansı açıklama oranı $(b:(p:o)) \times g$ deseninde diğer iki desende kestirilen toplam varyansı açıklama oranlarından yüksek çıkmıştır. $(b:(p:o)) \times g$ deseninde kestirilen okul ana etkisine ait varyans bileşeni diğer iki desende kestirilen varyans bileşenlerine göre daha yüksek kestirildiğinden, $(b:(p:o)) \times g$ desenine göre diğer iki desende okullarda ölçülen özellik bakımından farklılaşmanın az olduğu sonucuna ulaşılmıştır. Çalışma grubunu Anadolu lisesi ve Fen lisesi okullarında öğrenim gören öğrenciler oluşturduğundan Teog sınavı sonrası okullara öğrenci alım puanları ve öğrencilerin hazırbulunmuşlukları dikkate alındığında okullarda ölçülen özellik bakımından farklılaşma beklenen bir durumdur.

$b:o$ etkisine ait varyans bileşeninin toplam varyansı açıklama oranı $(b:o) \times g \times p$ deseninde, $(b:o) \times (g:p)$ deseninde kestirilen varyans bileşeninden daha yüksek çıkmıştır. Her iki desenden de elde edilen varyans bileşenleri için toplam varyansı açıklama oranlarına bakıldığında, okullarda performans görevlerini yanıtlayan öğrencilerin ölçülen özellik bakımından benzeşik olmadığı ve birey-okul ortak etkileşiminin farklılaştığı, öğrencilerin geometri becerilerinin bir okuldan diğerine değiştiği sonucuna ulaşılmıştır.

Görev ana etkisine ait varyans bileşeni $(b:o) \times g \times p$ deseninde $(b:(p:o)) \times g$ deseninden daha yüksek kestirilmiştir. Ancak her iki desende de kestirilen varyans bileşeni oranlarına bakıldığında, her öğrencinin cevapladığı geometri görevleri arasında farklılaşmanın olduğu sonucuna ulaşılmıştır. Elde edilen bu sonuç, performans görevlerinin farklı güçlüklerde olduğu ve zorluk-kolaylık düzeylerinin öğrenciden öğrenciye değiştiği şeklinde yorumlanmıştır.

Puanlayıcı ana etkisine ait varyans oranlarına bakıldığında $(b:o) \times (g:p)$ deseninde $(b:o) \times g \times p$ desenine göre çok yüksek bir değer kestirilmiştir. Bunun temel nedenlerinden birinin puanlayıcıların farklı görevleri $(g:p)$ puanlaması olabileceği sonucuna ulaşılmıştır. Bir başka ifade ile, $(b:o) \times (g:p)$ deseninde puanlayıcıların farklı görevleri puanlamasından dolayı puanlayıcıların görevlerle ilgili farklı yaşantılara sahip olmasından kaynaklanabileceği sonucuna ulaşılmıştır.

op ortak etkisine ait varyans değeri $(b:o) \times g \times p$ deseninde sıfıra yakın çıkarken $(b:o) \times (g:p)$ deseninde ise negatif değer aldığından dolayı sıfır olarak kabul edilmiştir. Varyansların sıfır ve sıfıra yakın çıkması, okul-puanlayıcı ortak etkisinin farklılaşmadığı, puanlayıcıların performanslarının okuldan okula farklılık göstermediği; diğer bir ifadeyle

puanlayıcıların farklı okullarda bulunan öğrencilerin geometri performanslarını puanlarken benzer cömertliğe/katılığa sahip olduğu sonucuna ulaşılmıştır.

og ortak etkisine ait varyans bileşenin toplam varyansı açıklama oranı $(b:o) \times g \times p$ ve $(b:(p:o)) \times g$ desenlerinde küçük kestirildiğinden, her iki desende de okul-görev ortak etkisinin farklılaşmadığı, geometri performans görevlerinin farklı okullarda bulunan bireyler tarafından aynı güçlükte algılandığı sonucuna ulaşılmıştır.

bp:o ortak etkisine ait varyans değeri $(b:o) \times (g:p)$ deseninde $(b:o) \times g \times p$ deseninden daha yüksek kestirilmiştir. Bu durumun $(b:o) \times (g:p)$ deseninde performans görevlerinin puanlayıcı yüzeyinde yuvalanmasından; diğer bir ifadeyle farklı puanlayıcıların farklı görevleri puanlamasından kaynaklı olarak yüksek çıktığı görülmektedir. Her iki desen için bp:o varyans bileşenin toplam varyansı açıklama oranlarına bakıldığında bu değer küçük olması, okulların ölçülen özellik bakımından benzeşik olduğu ve birey-puanlayıcı-okul ortak etkileşiminin farklılaşmadığı, öğrencilerin geometri becerilerinin bir okuldan diğerine değişmediği, puanlayıcıların her bir öğrencinin geometri performans görevlerini puanlarken bir okuldan diğerine tutarlı davrandığı sonucuna ulaşılmıştır. Bunun yanı sıra $(b:(p:o)) \times g$ deseninde b:p:o ortak etkisine ait varyans bileşenin toplam varyansı açıklama oranı diğer iki desendeki bp:o ortak etkisine ait varyans bileşenin toplam varyansı açıklama oranından yüksek çıkmıştır. Bu bulgu, farklı okullarda bulunan bireylerin geometri becerileri bakımından okuldan okula farklılaştığını, okulların ölçülen özellik bakımından benzeşik olmadığı ve birey-puanlayıcı-okul ortak etkileşiminin farklılaştığını, puanlayıcıların farklı okullarda bulunan her bir öğrencinin geometri performans görevlerini puanlarken puanlayıcının bir okuldan diğerine tutarlı davranmadığını göstermektedir. Bu bulgular dikkate alındığında, desenlerin yapısının varyans bileşenlerinin oranlarını, dolayısıyla yorumlanmasını farklılaştırdığı görülmüştür. Bir başka ifade ile desenlerde yuvalanmış veya çaprazlanmış yüzeylerin sayısı değiştiğinde varyans bileşenlerinin açıkladıkları varyans oranlarının da değiştiği sonucuna ulaşılmıştır.

Okul-puanlayıcı-görev ortak etkileşiminin açıkladığı varyans oranına bakıldığında da her üç desen için farklı sonuçlar elde edilmiştir. Bu durumda desenlerin farklılaşmasıyla varyans bileşenlerinin açıklama oranlarının değiştiğinin bir başka kanıtı olarak ortaya çıkmıştır.

Bütün bulgular dikkate alındığında, aynı veri seti üzerinden kurgulanan desenlerde varyans bileşenlerine ilişkin büyüklükler, varyans değişimlerinin büyüklük açısından sıralamadaki yeri farklılıklar göstermektedir. Bu durum, söz konusu bileşenlere ilişkin yorumlamaları da değiştirmektedir. Örneğin $(b:o) \times p \times g$ ve $(b:o) \times (g:p)$ desenlerinde okula ilişkin varyans değişiminin diğer bileşenlere göre küçük olması okullardan kaynaklı farklılığın az olduğu biçiminde yorumlanırken; $(b:(p:o)) \times g$ deseninde okullara ait varyans bileşenin çok yüksek olması, okullarda geometri becerileri bakımından farklılaşmanın fazla olduğu biçiminde yorumlanmıştır. Bunun yanı sıra söz konusu desenlerdeki artık varyans değerinin de değiştiği

görülmektedir. $(b:o)xp_xg$ deseni için kestirilen artık varyans bileşeni, $(b:o)x(g:p)$ ve $(b:(p:o))xg$ desenlerine göre oldukça küçük bulunmuştur. Bu bağlamda G kuramında duruma ilişkin kurgulanan desenin doğruluğunun yapılacak yorumların güvenilirliğini ve geçerliğini de etkileyeceği sonucuna ulaşılmıştır. Desenlerin çaprazlanmış ve/veya yuvalanmış olmasının da ana etkiler ile etkileşim etkilerinin varyans bileşenlerinin büyüklüklerinde farklılıklar yarattığı görülmektedir. Bir başka ifade ile, çaprazlanmış desenlerde ana etkilere ait varyans yüzdeleri çok büyük ya da çok küçük olabilirken yuvalanmış desenlerin açıkladığı varyans yüzdeleri değişebilmektedir. Bu durum Chang ve Hocevar (2000), Nalbantoğlu (2009), Alkan (2013), Nalbantoğlu ve Başusta (2015) ve Doğan ve Anadol (2017), tarafından yapılan çalışmalarla da desteklenmektedir.

$(b:o)xp_xg$, $(b:o)x(g:p)$ ve $(b:(p:o))xg$ desenlerinden elde edilen G ve Phi Katsayıları incelendiğinde 8 puanlayıcıdan her birinin farklı iki okulda bulunan 64 er öğrenciyi 8 görev doğrultusunda puanlamasıyla tasarlanan $(b:o)xp_xg$ desenine göre yapılan G çalışmasında G katsayısı 0.84, Phi katsayısı ise 0.78 olarak kestirilmiştir. Araştırmada 8 puanlayıcı içinde belirlenen 4 puanlayıcıdan her birinin farklı okullarda bulunan 64 öğrenciyi 2 'şer görev doğrultusunda puanlamasıyla tasarlanan $(b:o)x(g:p)$ desenine göre yapılan G çalışmasında G katsayısı 0.75, Phi katsayısı ise 0.56 olarak kestirilmiştir. Araştırmada farklı iki okulda bulunan 64 öğrenciyi 8 puanlayıcıdan her birinin 16 şarlı olarak ($8 \times 16 = 128$) 8 görev doğrultusunda puanlamasına ilişkin $(b:(p:o))xg$ desenine göre yapılan G çalışmasında G katsayısı 0.90, Phi katsayısı ise 0.87 olarak kestirilmiştir. Elde edilen bu katsayılar benzer durumda benzer nitelikteki puanlayıcıların bulunduğu durumlarda öğrencilerin geometri görevlerine ilişkin performanslarının genellenebilir nitelikte olduğunun bir göstergesidir.

Daha öncede belirtildiği gibi *evren puanı*, genelleme evrenindeki bütün koşulların ölçme objeleri için ortalama puanıdır. Bütün ölçme objelerinin evren puanlarının varyansı, *evren puan varyansı* olarak adlandırılır ve bu da KTK' deki gerçek puan varyansına benzerdir (Brennan ,1992). Bu araştırmada ölçme objesi öğrencilerdir. Her bir desen için bireye ait varyans bileşenleri farklı hesaplandığından, evren puan varyansları da farklıdır. Eşitlik 5 ve 6 'da belirtildiği gibi G ve Phi katsayıları evren puanına bağlı olarak hesaplandığından katsayılar farklılık göstermektedir. Çalışmamızda yer alan bütün desenlerin yuvalanmış olmasından dolayı bireylere ait varyans bileşenleri diğer değişkenlik kaynaklarından bağımsız olarak elde edilememekte ve yorumlanamamaktadır. Aynı zamanda desenlerin yapısı gereği içerdikleri varyans kaynakları sayısı da farklılık göstermektedir. G ve Phi katsayıları hesaplanırken paydada bulunan varyans bileşen sayısı da farklı olduğundan katsayılar da farklı büyüklüklerde bulunmuştur. Paydada bulunan varyans kaynak sayısı arttıkça katsayılar küçülecek; varyans kaynak sayısı azaldıkça da katsayılar büyüyecektir. Birinci desene ait G ve Phi katsayılarının ikinci desene göre yüksek olmasının sebebi; artık varyans değerinin ikinci desene göre oldukça

düşük olması; birinci desenin G ve Φ katsayılarının üçüncü desene göre düşük olması ise varyans bileşenlerinin sayıca fazla olması ile açıklanmıştır.

Bu çıkarımlar doğrultusunda üçüncü desenin varyans bileşenlerini açıklama oranı daha az gibi görünse de G ve Φ katsayıları diğer iki desene göre daha yüksek çıkmıştır. Bunun nedeni de desenin yapısındaki varyans bileşen sayısının az olması olarak açıklanmıştır. Bir desende ne kadar çok çaprazlanmış yüzey varsa ayrı ayrı elde edilen varyans bileşen sayısında o kadar artacağı görülmektedir. Ele alınan üç farklı desen ve her bir desenden elde edilen varyans bileşenlerinin sayısı farklı olduğundan elde edilen G ve Φ katsayılarında farklı olduğu sonucuna ulaşılmıştır.

4.1. Öneriler

1) Bu çalışmada geometri performansının ölçülmesinin güvenilirliği Genellebilirlik kuramında $(b:o)xp_xg$, $(b:o)x(g:p)$ ve $(b:(p:o))xg$ desenleri ile analiz edilmiştir. Benzer çalışmalar farklı desenler ve farklı değişkenlik kaynakları ele alınarak yapılabilir.

2) Genellebilirlik kuramının aynı desenleri ve aynı değişkenlik kaynakları ele alınarak dengelenmemiş desenlerde benzer çalışmalar yapılabilir.

3) Yapılacak performans ölçme çalışmalarında sadece Genellebilirlik kuramına dayalı güvenilirliğin belirlenmesi yerine diğer kuramlardan en az birinin de beraberinde kullanılması yararlı olabilir.

4) Genellebilirlik kuramında yer alan farklı karar çalışmalarıyla farklı sayıdaki görev ya da puanlayıcı sayılarına dayalı güvenilirlik tahminleri yapılarak sonraki ölçme çalışmaları için en uygun ve ekonomik puanlayıcı ve görev sayılarına önceden karar verilebilir.

5) Geometri performansının puanlanmasında hem bütünsel hem de analitik dereceli puanlama anahtarı kullanılarak aynı çalışma üzerinde bu iki puanlama yöntemi karşılaştırılarak benzer bir çalışma uygulanabilir.

6) Geometri dersi dışında farklı derslerde uygulanan performans ölçümleri için benzer güvenilirlik çalışmaları yapılabilir.

7) Geometrinin alt alanları (geometri-analitik geometri) değişkenlik kaynağı kabul edilerek yüzey sayısı artırılıp benzer çalışmalar yapılabilir.

KAYNAKLAR

- [1]. Aktaş, M., Başusta, B. ve Alıcı, D.(2016). *Farklı puanlama anahtarlarının güvenilirliğinin genellenabilirlik kuramı ile incelenmesi*. 1-3 Eylül 2016 Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi. Akdeniz Üniversitesi Eğitim Fakültesi, Antalya.
- [2]. Alharby, E. R. (2006). *A comparison between two scoring methods, holistic vs. Analytic using two measurement models, the generalizability theory and the many facet rasch measurement within the context of performance assessment*. Yayınlanmamış doktora tezi. The Pennsylvania State University.
- [3]. Alkan, M. (2013). *Pısa 2009 okuma becerileri açık uçlu sorularının puanlanmasında genellenebilirlik kuramındaki farklı desenlerin karşılaştırılması*. Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- [4]. Anıl, D., Büyükkıdık, S. (2012). Genellenebilirlik kuramında dört facetli karışık desen kullanımı için örnek bir uygulama. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, Kış, 3(2)*, 291-296.
- [5]. Atılğan, H. (2004). *Genellenebilirlik kuramı ve çok değişkenlik kaynaklı Rasch modelinin karşılaştırılmasına ilişkin bir araştırma*. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Ankara.
- [6]. Atılğan, H. ve Tezbaşaran, A. A. (2005). Genellenebilirlik kuramı alternatif karar çalışmaları ile senaryolar ve gerçek durumlar için elde edilen g ve phi katsayılarının tutarlılığının incelenmesi. *Eurasian Journal of Educational Research, 18*, 236-252.
- [7]. Atılğan, H. (Ed.). (2009). *Eğitimde ölçme ve değerlendirme*. Ankara: Anı Yayıncılık.
- [8]. Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- [9]. Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: PegemA Yayıncılık.
- [10]. Brennan, R. L. (1992). *Elements of generalizability theory (rev. ed)*. Iowa City, IA. American College Testing.
- [11]. Brennan, R. L. (2001). *Generalizability theory*. New York: Springer- Verlog.
- [12]. Brennan, R. L. (2001). Variability of estimated variance components and related statistics in a performance assessment. *Applied measurement in education, 14(2)*, 191-203.
- [13]. Brennan, R. L. (2011). Generalizability theory and classical Test theory. *Center for advanced studies in measurement and assessment. Applied measurement in education, 24*: 1-21.
- [14]. Brualdi, A. (1998). *Implementing performance assessment in the classroom*. Practical Assessment, Research & Evaluation, 6(2), <http://PAREonline.net/getvn.asp?v=6&n=2>.
- [15]. Büyükkıdık, S. (2012). *Problem çözme becerisinin değerlendirilmesinde puanlayıcılar arası güvenilirliğin Klasik Test Kuramı ve genellenebilirlik kuramına göre karşılaştırılması*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara.
- [16]. Büyüköztürk, Ş. (2007). Performansa dayalı durum belirleme nedir? *İlköğretmen Dergisi, 8*, 28- 32.
- [17]. Chang, L. & Dennis, H. (2010). Models of generalizability theory in analyzing existing faculty evaluation data . *Applied Measurement In Education, 13(3)*, 255-275.
- [18]. Chien, Y.(2008). *An investigation of testlet-based item response models with a random facets design in generalizability theory*. Doktora Tezi, The University of Iowa.
- [19]. Cohen, R. J. & Swerdlik, M. E. (2013). *Psikolojik test ve değerlendirme* (Çev. E. Tavşancıl). Nobel yayıncılık (7. Basım).
- [20]. Crocker, L., & Algina, J.(1986). *Introduction to classical and modern test theory*. Harcourt Brace Javanovich College Publishers, USA.
- [21]. Çakıcı, D. (2011). *Genellenebilirlik kuramı ve lojistik regresyona dayalı hesaplanan puanlayıcılar arası tutarlığın karşılaştırılması*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara.

- [22].Doğan, C.D ve Anadol, H.Ö. (2017). Genellenebilirlik kuramında tümüyle çaprazlanmış ve maddelerin puanlayıcılara yuvalandığı desenlerin karşılaştırılması. *Kastamonu Eğitim Dergisi Cilt:25 No: 1*, 361-372.
- [23].Erkuş, A. (2003). *Psikometri üzerine yazılar*. Ankara: TPD Yayınları.
- [24].Fan, X. & Sun, S. (2013). *Generalizability theory as a unifying framework of measurement reliability in adolescent research*. Journal of early adolescence 2014, vol 34(1), 38 –65
- [25].Güler, N. (2008). *Klasik test kuramı, genellenebilirlik kuramı ve Rasch modeli üzerine bir araştırma*. Doktora Tezi. Hacettepe Üniversitesi, Ankara.
- [26].Güler, N, Kaya Uyanık ve Taşdelen Teker, G. (2012). *Genellenebilirlik kuramı*. PegemA Yayıncılık.
- [27].Karasar, N. (1998). *Araştırmalarda rapor hazırlama yöntemi*. Ankara: Pars Matbaacılık
- [28].Kim,Y. (1993). *The use of a 2-parameter graded IRT model and generalizability theory in the development and validation of a mathematics performance test*. Doktora tezi, University of Pittsburgh.
- [29].Moskal, Barbara M. & JonA. Leydens (2000). Scoring rubric development: Validity and reliability. *Practical Assessment, Research & Evaluation*, 7(10). Available online: <http://PAREonline.net/getvn.asp?v=7&n=10>
- [30].Mushquash, C. & O'Connor, B. P. (2006). Spss and sas programs for generalizability theory analysis. *Behavior Research Methods*. 38 (3), 542-547.
- [31].Nalbantoğlu, F. (2009). *Performans ölçümlerinde genellenebilirlik kuramıyla farklı desenlerin karşılaştırılması*. Yayınlanmamış Yüksek Lisans Tezi. Hacettepe Üniversitesi, Ankara.
- [32]. Nalbantoğlu, F. ve Başusta, B. (2015). *Genellenebilirlik kuramıyla dikiş atma ve alma becerileri istasyonu güvenirlüğünün değerlendirilmesi*. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 6 (1), 107-116.
- [33].Özberk, E.H. (2012). *Genellenebilirlik kuramı karar çalışmalarında kullanılan farklı katsayıların karşılaştırılması*. Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara.
- [34].Shavelson, J. R, & Webb N. M. (1991). *Generalizability theory :a primer*. Newbury Park. CA: Sage Publications.
- [35].Taylor, M. A. & Pastor, D. A. (2013). An application of generalizability theory to evaluate the technical quality of an alternate assessment. *Applied Measurement in Education*, 26, 279–297.
- [36].Tekin, H. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara : Yargı Yayınevi.
- [37].Turgut, M.F., Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. PegemA Yayıncılık.
- [38].Yin, Y. & Shevalson, R. (2004). *Application of generalizability theory to concept-map assessment research*. 26 Mayıs 2016 tarihinde <http://www.stanford.edu/dept/SUSE/SEAL> adresinden erişildi.
- [39].Yin, Y. (2008). Application of generalizability theory to concept map assessment research. *Applied Measurement In Education*, 21, 273–291.

EKLER

EK 1

Soru Kağıdı

Sevgili Öğrenciler,

Bu çalışma, bilimsel bir projenin verilerini elde etmek üzere yapılmaktadır. Sizden istenen, aşağıda yer alan soruları cevaplamanızdır. Verdiğiniz cevaplar, araştırmanın sonuçları açısından bilimsel önem taşıdığından cevapları eksiksiz, ayrıntılı ve aşamalı olarak yazmanızı rica eder, katkılarınız için teşekkür ederim.

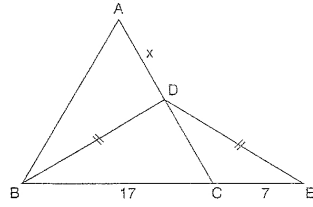
Ad :

Okul :

Soyad :

Alan :

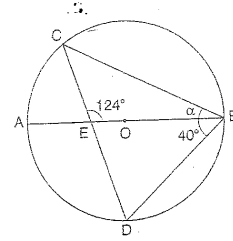
1.



ABC eşkenar üçgen
DBE ikizkenar üçgen
 $|BD| = |DE|$
 $|BC| = 17$ br
 $|CE| = 7$ br olduğuna göre,
 $|AD| = x$ kaç br dir?

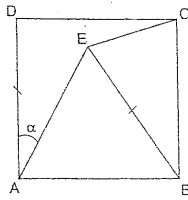
3.

$[AB]$ çap, O merkez
 $[AB] \cap [CD] = \{E\}$
 $m(\widehat{BEC}) = 124^\circ$
 $m(\widehat{DBA}) = 40^\circ$



Buna göre, $m(\widehat{ABC}) = \alpha$ kaç derecedir?

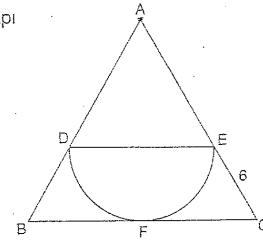
2.



ABCD kare
 $2m(\widehat{EAB}) = 3m(\widehat{ECB})$
 $|AD| = |EB|$ olduğuna göre,
 $m(\widehat{DAE}) = \alpha$ kaç derecedir?

4.

ABC bir üçgen
 $[DE]$ çemberin çapı
 $[DE] \parallel [BC]$
 $3|BD| = |AD|$
 $|AC| = |BC|$
 $|EC| = 6$ cm



Yukarıdaki verilere göre, çemberin yarıçapı kaç cm dir?

Soru Kağıdı (devamı)

5.

Analitik düzlemde denklemi;

$$(k + 3)x + (k - 1)y + 2k + 10 = 0$$

olan doğruların geçtiği noktadan ve orjinden geçen doğrunun denklemi aşağıdakilerden hangisidir?

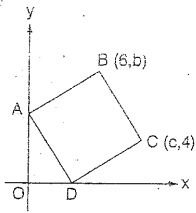
7.

Ayrıtları a, b, c olan bir dikdörtgenler prizmasında

$$\frac{1}{a} + \frac{1}{b} + \frac{1}{c} = \frac{9}{16}$$

ve prizmanın alanı 360 br^2 olduğuna göre, hacmi kaç br^3 tür?

6.



Dik koordinat sistemindeki ABCD bir kare

B(6,b)

C(c,4)

Yukarıdaki verilere göre, A(ABCD) kaç br^2 dir?

8.

Yüksekliği 15 cm olan bir piramit tabanına paralel bir düzlemle kesiliyor.

Üstteki küçük piramitin hacminin kesik piramitin hacmine oranı $\frac{27}{98}$ olduğuna göre, kesik piramitin yüksekliği kaç cm dir?

EK 2

Geometri Performansının Ölçülmesine İlişkin Bütünsel Dereceli Puanlama Anahtarı

Puan	Ölçütler	Tanımlama
4	Örnek cevap	-Soru tam cevaplandırılmıştır. -Sorunun cevabında hiç hata yoktur. -Sorunun çözülmesi için gerekli matematiksel bilgisi vardır.
3	Yeterli Cevap	-Soruya verilen cevap tam değildir. -Sorunun çözülmesi için gerekli bilgisi vardır ancak bazı basit matematiksel hatalar vardır.
2	Kısmen Yeterli Cevap	-Soruya verilen cevap çok az tatmin edicidir. -Soruya uygun başlar ancak cevabın geri kalanı yanlıştır. -Sorunun çözülmesi için gerekli çok az bilgisi vardır.
1	Yetersiz Cevap	-Soruya verilen cevap baştan sona yanlıştır. -Sorunun çözülmesi için gerekli bilgisi hiç yoktur.
0	Cevap yok	- Soru cevaplandırılmamıştır.

EK 3

Puanlama Tablosu

	G1	G2	G3	G4	G5	G6	G7	G8
01B1								
01B2								
01B3								
01B4								
01B5								
01B6								
01B7								
02B1								
02B2								
02B3								
02B4								
02B5								
02B6								
02B7								

EK 4

File C:\Users\Samsung\Desktop\11.11.16\BOXPXG.gen - [2016-11-11 11:11]

BOXPXG

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
O	O	2	INF	
B:O	B:O	64	INF	
P	P	8	INF	
G	G	8	INF	

Analysis of variance

Source	SS	df	MS	Components				
				Random	Mixed	Corrected	%	SE
O	730.93372	1	730.93372	0.16445	0.16445	0.16445	9.8	0.14576
B:O	2845.31226	126	22.58184	0.27939	0.27939	0.27939	16.6	0.04423
P	857.05945	7	122.43706	0.10976	0.10976	0.10976	6.5	0.05645
G	2133.90125	7	304.84304	0.26182	0.26182	0.26182	15.5	0.14118
OP	46.63757	7	6.66251	0.01151	0.01151	0.01151	0.7	0.00614
OG	233.55750	7	33.36536	0.05564	0.05564	0.05564	3.3	0.03072
BP:O	355.03735	882	0.40254	0.02409	0.02409	0.02409	1.4	0.00244
BG:O	3976.02563	882	4.50797	0.53727	0.53727	0.53727	31.9	0.02681
PG	193.79309	49	3.95496	0.02639	0.02639	0.02639	1.6	0.00618
OPG	28.30090	49	0.57757	0.00575	0.00575	0.00575	0.3	0.00179
BPG:O	1295.54663	6174	0.20984	0.20984	0.20984	0.20984	12.4	0.00378
Total	12696.10535	8191					100%	

G Study Table
(Measurement design BO/PG)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
O	0.16445		
B:O	0.27939		
	P		0.01372	10.7
	G		0.03273	25.4
	OP	0.00144	1.8	0.00144	1.1
	OG	0.00696	8.5	0.00696	5.4
	BP:O	0.00301	3.7	0.00301	2.3
	BG:O	0.06716	82.0	0.06716	52.1
	PG		0.00041	0.3
	OPG	0.00009	0.1	0.00009	0.1
	BPG:O	0.00328	4.0	0.00328	2.5
Sum of variances	0.44385		0.08193	100%	0.12879	100%
Standard deviation	0.66622		Relative SE: 0.28624		Absolute SE: 0.35888	
Coef_G relative	0.84					
Coef_G absolute	0.78					

Grand mean for levels used: 2.69324

Variance error of the mean for levels used: 0.13609

Standard error of the grand mean: 0.36890

File C:\Users\Samsung\Desktop\11.11.16\BOXGP.gen - [2016-11-28 10:31]

BOXGP

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
O	O	2	INF	
B:O	B:O	64	INF	
P	P	4	INF	
G:P	G:P	2	INF	

Analysis of variance

Source	SS	df	MS	Components				
				Random	Mixed	Corrected	%	SE
O	72.25000	1	72.25000	0.13494	0.13494	0.13494	6.9	0.11523
B:O	334.50000	126	2.65476	0.22390	0.22390	0.22390	11.4	0.04221
P	495.47656	3	165.15885	0.52391	0.52391	0.52391	26.6	0.41619
G:P	143.21094	4	35.80273	0.23181	0.23181	0.23181	11.8	0.16384
OP	4.10156	3	1.36719	-0.03870	-0.03870	-0.03870	0.0	0.02847
OG:P	24.52344	4	6.13086	0.08528	0.08528	0.08528	4.3	0.05531
BP:O	326.42188	378	0.86355	0.09520	0.09520	0.09520	4.8	0.03780
BG:OP	339.26563	504	0.67315	0.67315	0.67315	0.67315	34.2	0.04232
Total	1739.75000	1023					100%	

G Study Table
(Measurement design BO/GP)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
O	0.13494		
B:O	0.22390		
	P		0.13098	47.0
	G:P		0.02898	10.4
	OP	(0.00000)	0.0	(0.00000)	0.0
	OG:P	0.01066	9.0	0.01066	3.8
	BP:O	0.02380	20.1	0.02380	8.5
	BG:OP	0.08414	70.9	0.08414	30.2
Sum of variances	0.35885		0.11860	100%	0.27856	100%
Standard deviation	0.59904		Relative SE: 0.34439		Absolute SE: 0.52778	
Coef_G relative	0.75					
Coef_G absolute	0.56					

Grand mean for levels used: 2.57813
 Variance error of the mean for levels used: 0.23535
 Standard error of the grand mean: 0.48513

File C:\Users\Samsung\Desktop\BPO.gen - [2017-06-12 15:14]

BPOXG

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
O	O	2	INF	
P:O	P:O	4	INF	
B:P:O	B:P:O	16	INF	
G	G	8	INF	

Analysis of variance

Source	SS	df	MS	Components				SE
				Random	Mixed	Corrected	%	
O	312.84766	1	312.84766	0.58821	0.58821	0.58821	30.6	0.49902
P:O	64.68359	6	10.78060	0.05624	0.05624	0.05624	2.9	0.04229
B:P:O	311.71875	120	2.59766	0.23609	0.23609	0.23609	12.3	0.04180
G	250.95313	7	35.85045	0.25980	0.25980	0.25980	13.5	0.13238
OG	18.16797	7	2.59542	0.01410	0.01410	0.01410	0.7	0.01993
PG:O	71.09766	42	1.69280	0.06149	0.06149	0.06149	3.2	0.02266
BG:P:O	595.53125	840	0.70897	0.70897	0.70897	0.70897	36.8	0.03455
Total	1625.00000	1023					100	
	0						%	

G Study Table
(Measurement design BPO/G)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
O	0.58821		
P:O	0.05624		
B:P:O	0.23609		
	G		0.03248	24.9
	OG	0.00176	1.8	0.00176	1.4
	PG:O	0.00769	7.8	0.00769	5.9
	BG:P:O	0.08862	90.4	0.08862	67.9
Sum of variances	0.88054		0.09807	100%	0.13055	100%
Standard deviation	0.93837		Relative SE: 0.31316		Absolute SE: 0.36131	
Coef_G relative	0.90					
Coef_G absolute	0.87					

Grand mean for levels used: 2.65625

Variance error of the mean for levels used: 0.33799

Standard error of the grand mean: 0.58137

ÖZGEÇMİŞ

Adı ve Soyadı: Zarife TAŞTAN

Doğum Tarihi : 01 EKİM 1984

E-mail : zariphe@mynet.com

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	Matematik	Ege Üniversitesi	2004-2008
Tezsiz Yüksek Lisans	Matematik Öğretmenliği	Mersin Üniveristesi	2008-2009

Tarih : 07.2017