

**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI**

**PERSONEL ALIMINDA KULLANILAN BİR YAZILI SINAV
SONUCUNUN GENELLENEBİLİRLİK KURAMINDAKİ FARKLI
DESENLERLE KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

İREM DEMİRBİLEK ZORBA

**ANKARA
ARALIK, 2020**



**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI**

**PERSONEL ALIMINDA KULLANILAN BİR YAZILI SINAV
SONUCUNUN GENELLENEBİLİRLİK KURAMINDAKİ FARKLI
DESENLERLE KARŞILAŞTIRILMASI**

YÜKSEK LİSANS TEZİ

İREM DEMİRBİLEK ZORBA

DANIŞMAN: DOÇ. DR. KAAAN ZÜLFİKAR DENİZ

**ANKARA
ARALIK, 2020**

ETİK İLKELERE UYGUNLUK

Tez içindeki bütün bilgileri akademik yazım kurallarına uygun biçimde raporlaştırdığımı ve bunları etik ilkelere (atıfta bulunulan tüm yapıtlara kaynaklarda yer verilmesi, tezde kullanılan bilgi ve belgelere resmi yollarla ulaşılması ve bunların aslı bozulmadan kullanılması vb.) uygun olarak elde ettiğimi ve sunduğumu bildiririm.


İrem DEMİRBILEK ZORBA

ÖZET

PERSONEL ALIMINDA KULLANILAN BİR YAZILI SINAV SONUCUNUN GENELLENEBİLİRLİK KURAMINDAKİ FARKLI DESENLERLE KARŞILAŞTIRILMASI

DEMİRBILEK ZORBA, İrem

Yüksek Lisans Tezi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı

Tez Danışmanı: Doç. Dr. Kaan Zülfiyar DENİZ

Aralık, 2020, xiii + 69 sayfa

Bu araştırmanın amacı, açık uçlu maddelerden oluşmuş personel alımı sınavına katılan bireylerin cevaplarının, farklı puanlayıcılar tarafından değerlendirilmesiyle elde edilen verilere genellenebilirlik kuramındaki farklı desenler uygulanarak G ve K çalışmalarının yapılması ve bu çalışmaların sonucunda elde edilen G ve Phi katsayılarının analiz edilip yorumlanmasıdır. Araştırmanın çalışma grubu, ASYM'nin uyguladığı açık uçlu maddelerden oluşan personel alımı sınavına katılan 211 bireyin Hukuk alt testinde yer alan maddelerin iki puanlayıcı ile değerlendirilmesi ile elde edilen verilerden oluşmuştur. Testte bireylere dört soru sorulmuştur ve bu sorulardan istedikleri iki tanesini cevaplamaları istenmiştir. Bireylerin cevapları da iki puanlayıcı tarafından değerlendirilmiştir. Bireylerin en çok seçtikleri maddeler belirlenerek veri sınırlandırılmış ve 93 birey ile oluşturulan veri üzerine (bxm_{xp}) çaprazlanmış deseni uygulanmıştır. Bireylerin maddeler ve puanlayıcılarla çapraz tasarlandığı ve maddelerin puanlayıcılara yuvalandığı (bx(m:p)) yuvalanmış deseni uygulamak amacıyla veriler sınırlandırılmış ve 40 bireyden elde edilen veri üzerinde yuvalanmış desen uygulanmıştır. Araştırmanın alt amaçlarını analiz edebilmek amacıyla çaprazlanmış ve yuvalanmış desenleri ile G çalışması yaparak göreceli ve mutlak hata varyansları ve G ve Phi katsayıları hesaplanmıştır ve K çalışması yapılmıştır. Desenlere ait varyans bileşenlerinin tahmin edilmesi, bileşenlerin varyans açıklama yüzdelerinin hesaplanması ve K çalışmalarının yapılması amacıyla 2006 yılında Jean Cardinet tarafından geliştirilen EduG programı kullanılmıştır.

G çalışması sonuçlarına göre (bxm_{xp}) deseninde G ve Phi katsayıları sırasıyla

0,33 ve 0,29 elde edilirken (bx(m:p)) deseninde G ve Phi katsayıları sırasıyla 0,76 ve 0,64 elde edilmiştir. K çalışması sonuçlarına göre (bxmxp) deseninde puanlayıcı sayısının artırılması ile (bx(m:p)) deseninde ise madde sayısının artırılması ile güvenilirliğin arttığı gözlemlenmiştir.

Anahtar Kelimeler: Genellenebilirlik Kuramı, Çaprazlanmış Desen, Yuvalanmış Desen, G çalışması, Karar çalışması.



ABSTRACT

A COMPARISON OF THE RESULT OF A WRITTEN EXAM USED IN PERSONNEL RECRUITMENT WITH DIFFERENT PATTERNS IN THE GENERALIZABILITY THEORY

DEMİRBILEK ZORBA, İrem

Master, Department of Measurement and Evaluation in Education

Supervisor: Doç. Dr. Kaan Zülfikar DENİZ

December, 2020, xiii + 69 pages

The purpose of this study is to conduct G and D studies by applying different designs in generalizability theory to the data obtained by evaluating the answers of individuals who participate in a job entrance exam consisting of open-ended items by different raters, and to analyze and interpret the G and Phi coefficients obtained as a result of these studies.

The study group of the study was obtained with the group of 211 individuals participating in the open-ended items applied by ASYM with two raters in the Law subtest. In the test, they were asked to select to answer two of four questions. Individuals' answers were also evaluated by two raters. The data was limited by determining the items that individuals chose the most and a crossed design (bxm_{xp}) was applied on the data created with 93 individuals. Individuals are cross-designed with items and raters and items are nested to raters in order to apply (bx(m:p)) the nested design, the data were limited and nested design was applied on the data obtained from 40 individuals. In order to analyze the sub-objectives of the study, the relative and absolute error variances and G and Phi coefficients were calculated by making G study with crossed and nested designs and D study was conducted. EduG program developed by Jean Cardinet in 2006 was used in order to estimate the variance components of the designs, calculate the variance explanation percentages of the components and conduct D studies.

According to the results of the G study, the G and Phi coefficients calculated 0,33 and 0,29 for (bxm_{xp}) design, and 0,76 and 0,64 for (bx(m:p)) design respectively. According to the results of the K study, it was observed that increasing the number of

raters in the (bxm_xp) design and increasing the number of items in the design (bx (m: p)) increased the reliability.

Keywords: Generalizability Theory, Crossed Design, Nested Design, G study, Decision study.



ÖNSÖZ

Eğitimde yapılan ölçme işlemleri dolaylı ölçme türüne örnektir. Bir ölçme aracı kullanılarak yapılan ölçme işlemine mutlaka belirli oranlarda hata karışır. Bu hataların en iyi derecede tespit edilip yorumlanabilmesi ölçme ve değerlendirme bilim dalının vazgeçilmez bir ayağıdır. Bu hataların tespit edilebilmesi için birçok kuram vardır. Bu kuramlardan bir tanesi de tezin çalışma alanı da olan Genellenebilirlik Kuramıdır.

Genellenebilirlik Kuramı ile ölçme süreçlerine karışan birden fazla hatanın aynı anda analiz edilebilmesine olanak sağlar. Genellenebilirlik Kuramında genellenebilirlik çalışması ve karar çalışması isimleriyle iki aşamalı bir çalışma yürütülür. Genellenebilirlik çalışmasında, analizde kullanılan değişkenlik kaynaklarının ölçme işleminde ne kadar varyans oluşturduğu ve bu varyansın toplam varyansa ne kadar etki ettiği elde edilebilmektedir. Aynı zamanda bağıl ve mutlak hata varyanslarına dayalı olarak elde edilen G ve Phi katsayıları ölçme işleminin güvenilirliği ile ilgili yorumlar yapılmasına olanak sağlamaktadır. Karar çalışması ile değişkenlik kaynaklarının sayılarının değiştirilmesi senaryoları kurularak hangi değişikliklerin daha güvenilir sonuçlar elde edilmesini sağlayacağı belirlenmeye çalışılmaktadır. Böylece analizde kullanılan değişkenlik kaynaklarının hangi düzeylerden daha güvenilir sonuçlar elde edildiğini yorumlanabilmesine olanak sağlanmaktadır.

Bu araştırmada, personel alımı sınavından elde edilen veriler üzerine G kuramı uygulanarak genellenebilirlik çalışması ve karar çalışması uygulanmıştır. Elde edilen değerler karşılaştırılıp yorumlanmıştır. En güvenilir sonuçları elde edebilmek için olması gereken değişiklikler belirlenmeye çalışılmıştır.

TEŞEKKÜR

Tez çalışmam sürecinde değerli vakitlerini ayırarak her türlü sorumu yanıtlayan ve yol gösteren tez danışmanım ve kıymetli hocam Sayın Doç. Dr. Kaan Zülfikar DENİZ'e bana ve çalışmama olan katkıları için teşekkürlerimi sunarım.

Tez savunma jürimde yer alarak verdikleri geri dönütler ile çalışmama katkı sağlayan değerli hocalarım Sayın Prof. Dr. Selahattin GELBAL ve Doç. Dr. Celal Deha DOĞAN'a teşekkürlerimi sunarım.

Yüksek lisans eğitimim boyunca derslerini alarak kendilerinden çok şey öğrendiğim değerli hocalarım Sayın Prof. Dr. Nizamettin KOÇ'a, Sayın Prof. Dr. Ezel TAVŞANCIL'a, Sayın Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU'na, Sayın Doç. Dr. Deniz GÜLLEROĞLU'na ve Sayın Dr. Ömer KUTLU'ya verdikleri emeklerden ötürü çok teşekkür ederim.

Tezimi dil, anlatım ve düzen açısından inceleyen, yaşadığım her zorlukta yanımda olan sevgili eşim Ömer Özen ZORBA'ya teşekkürleri sunarım. Ayrıca tez çalışmam süresince bu kadar anlayışlı olan biricik oğlum Kerem ZORBA'ya teşekkür ederim.

Hayatım boyunca beni destekleyen, tüm başarı ve başarısızlıklarımda yanımda olan kıymetli ailem annem, babam ve ablama teşekkürlerimi sunarım.

İÇİNDEKİLER

	Sayfa
ETİK İLKELERE UYGUNLUK	iii
ÖZET	iv
ABSTRACT	vi
ÖNSÖZ.....	viii
TEŞEKKÜR.....	ix
İÇİNDEKİLER.....	x
TABLolar DİZİNİ	xii
ŞEKİLLER DİZİNİ.....	xii
KISALTMALAR	xiii
BÖLÜM 1	1
GİRİŞ	1
Problem Durumu	1
Klasik Test Kuramı	1
Klasik Test Kuramında Güvenirlik ve Güvenirliğin Kestirilme Yöntemleri.....	3
Klasik Test Kuramında Kestirilen Güvenirliklerin Sınırlılıkları	7
Genellenebilirlik Kuramı	8
G Kuramı ve KTK'nin Sınırlı ve Güçlü Yönlerinin Karşılaştırılması.....	9
Kabul Edilebilir Gözlemler ve G Çalışması	11
Sonsuz Genelleme Evreni ve Karar (K) Çalışması.....	12
Çaprazlanmış ve Yuvalanmış Desenler	13
Hata Varyansları	18
Genellenebilirlik Katsayısı ve Güvenirlik Katsayısı	18
Açık Uçlu Maddeler.....	20
Amaç	22
Alt Amaçlar.....	22
Önem	23
Sayıtlar	24
Sınırlılıklar	24
İlgili Araştırmalar	24
Yurtiçinde Yapılan Araştırmalar	24
Yurtdışında Yapılan Araştırmalar.....	29
BÖLÜM 2	35
YÖNTEM.....	35
Araştırmanın Modeli	35
Çalışma Grubu.....	35
Verilerin Toplanması.....	35
Verilerin Çözümlemesi	36

BÖLÜM 3	37
BULGULAR VE YORUMLAR.....	37
Birey (b), Madde (m) Ve Puanlayıcı (p) Değişkenlerinin Çapraz Tasarlandığı (bxm _x p) Deseninin G Kuramı Sonuçları	37
(bxm _x p) Deseninin G Çalışması Sonucu Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri	37
(bxm _x p) Deseninde Madde ve Puanlayıcı Sayılarının Artırılması ve Azaltılmasıyla Yapılan K Çalışması.....	40
Birey (b) değişkenin çapraz, Madde (m) Ve Puanlayıcı (p) Değişkenlerinin Yuvalanmış Tasarlandığı (bx(m:p)) Deseninin G Kuramı Sonuçları	43
(bx(m:p)) Deseninin G Çalışması Sonucu Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri	43
(bx(m:p)) Deseninde Madde ve Puanlayıcı Sayılarının Artırılması ve Azaltılmasıyla Yapılan K Çalışması.....	46
(bxm _x p) ve (bx(m:p)) Desenleri ile Kestirilen G Çalışması Sonuçlarının Karşılaştırılması	49
(bxm _x p) ve (bx(m:p)) Desenleri ile Madde ve Puanlayıcı Sayılarının Değiştirilmesiyle Kestirilen Karar Çalışmaları Sonuçlarının Karşılaştırılması ..	50
BÖLÜM 4	55
SONUÇ VE ÖNERİLER	55
Sonuçlar.....	55
Birinci Alt Amaca İlişkin Sonuçlar.....	55
İkinci Alt Amaca İlişkin Sonuçlar	56
Üçüncü Alt Amaca İlişkin Sonuçlar	57
Dördüncü Alt Amaca İlişkin Sonuçlar.....	58
Öneriler.....	59
Uygulayıcılara Yönelik Öneriler.....	59
Araştırmacılara Yönelik Öneriler	60
KAYNAKLAR.....	61
EK-1: ASYM'den alınan veri kullanımı için izin belgesi.....	67
BENZERLİK BİLDİRİMİ	68
ÖZGEÇMİŞ	69

TABLolar DİZİNİ

Tablo	Sayfa
Tablo 1. İki yüzeyli evrenlerde varyans bileşenleri	16
Tablo 2. Beklenen ortalama kareler ve $p \times t \times r$ için tahmin edilen varyans bileşenleri	17
Tablo 3. Desende Kullanılan Yüzeylerin Sayısal Değerleri	37
Tablo 4. (bxm _{xp}) Deseni Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri	38
Tablo 5. (bxm _{xp}) Desenine Ait G ve Phi Katsayı Değerleri	40
Tablo 6. (bxm _{xp}) Deseni Madde ve Puanlayıcı Sayılarının Değiştirilmesiyle Kestirilen G ve Phi Katsayı Değerleri, Bağlı ve Mutlak Hata Varyansları.....	40
Tablo 7. Desende Kullanılan Yüzeylerin Sayısal Değerleri	44
Tablo 8. (bx(m:p)) Deseni Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri	44
Tablo 9. (bx(m:p)) Desenine Ait G ve Phi Katsayı Değerleri.....	46
Tablo 10. (bx(m:p)) Deseni Madde ve Puanlayıcı Sayılarının Değiştirilmesiyle Kestirilen G ve Phi Katsayı Değerleri , Bağlı ve Mutlak Hata Varyansları.....	47
Tablo 11. (bxm _{xp}) ve (bx(m:p)) Desenlerine Ait Varyans ve Toplam Varyans Açıklama Yüzdeleri.....	49
Tablo 12. (bxm _{xp}) ve (bx(m:p)) Desenleri için Madde ve Puanlayıcı Sayılarının Değiştirilmesiyle Oluşturulmuş Senaryolara Göre Elde Edilen Bağlı ve Mutlak Hata Varyansları.....	51
Tablo 13. (bxm _{xp}) ve (bx(m:p)) Desenlerinin K Çalışması Sonuçlarına göre G ve Phi Katsayı Değerleri	52

ŞEKİLLER DİZİNİ

Şekil	Sayfa
Şekil 1. Tek Yüzeyli Çaprazlanmış ve Yuvalanmış Desen Venn Şeması	14

KISALTMALAR

ASYM	Ankara Üniversitesi Sınav Yönetim Merkezi
KTK	Klasik Test Kuramı
MTK	Madde Tepki Kuramı
G Kuramı	Genellenebilirlik Kuramı
G Çalışması	Genellenebilirlik Çalışması
K Çalışması	Karar Çalışması
G Katsayısı	Genellenebilirlik Katsayısı
Phi Katsayısı	Güvenirlilik Katsayısı
“x”	Çaprazlama
“.”	Yuvalama
b	Birey
p	Puanlayıcı
m	Madde

BÖLÜM 1

GİRİŞ

Birinci bölümde, araştırmanın konusu olan problem, araştırmanın önemi, amacı ve alt amaçları, sayıltıları, sınırlılıkları ve araştırmada bulunan kısaltmalar bulunmaktadır.

Problem Durumu

Ölçme ve değerlendirme bilimi, temellerini var olan farkı bulmaktan alan gözlenen değerlerden yola çıkarak gözlenemeyen değerleri tahmin etme, bir başka ifade ile gözlenebilen değişkenlerden, doğrudan gözlenemeyen değişkenlerin elde edilmesini konu almaktadır. Bu işlemi yaparken de çeşitli kurallar ve varsayımlardan yola çıkarak elde etmeye çalıştığı değişkenlere ait ölçme sonuçlarını güvenilir, geçerli ve doğru bir şekilde gözlemlemeyi amaçlar. Eğitim ve psikolojide ölçülen özellikler ve bu özelliklerin ölçülmesi sonucu elde edilen sonuçlar da ölçülen özellikler gibi değişkendir. Bu nedenle eğitim ve psikolojide yapılan ölçmelerde her zaman bir miktar hata vardır. Bu yüzden, ölçme ve değerlendirme çalışmalarının amacı, bu hata miktarını kontrol altına alarak en aza indirmeye çalışmaktır. Bunun için de çeşitli varsayımlara ve kurallara dayalı ölçme kuramları ortaya atılmıştır. Bütün kuramların temelinde, güvenilir, geçerli, doğru ve tutarlı ölçümler yaparak hatalardan arınık sonuçlar elde etmeye çalışmak vardır. Bu kuramlardan en yaygın olarak tercih edileni Klasik Test Kuramıdır (Hambleton, 2004).

Klasik Test Kuramı

Eğitim ve psikolojideki işe vuruk olmayan tanımlamalar çeşitli sayı ve sembollerle ifade edilerek işe vuruk değişkenlere dönüştürülür. Klasik test kuramı (KTK), işe vuruk olarak tanımlanamayan eğitim ve psikolojideki bazı değişkenleri, ölçme aracılığı ile bilimsel kuramlarla bir araya getirmeye ve bunları açıklamaya yardımcı olur (Lord ve Novick,1968). KTK'nin açıklamak durumunda olduğu problemin çözümü için ortaya koyduğu bazı temel kavramlar ve sayıltılar vardır. Bu kavramlar gerçek puan, gözlenen puan ve hata kavramıdır.

Herhangi bir ölçme işleminde, ölçülen özelliğin ölçme süresince değişmeyen durumunu gösteren, bu özelliğin gerçek değerine karşı geldiği varsayılan sayıya gerçek puan denir. Bu hipotetik bir kavramdır çünkü çoğu zaman bir niteliğin o varlıktaki gerçek miktarına ulaşmak mümkün değildir. Bu kullanılan test tekniğinden, ölçmeyi yapan bireyden, ölçülen varlıktan ve ölçülen özelliğin doğasından kaynaklanan sorunlardan kaynaklı olabilir. Eğitimde ve psikolojide, ölçtüğümüz özelliği doğrudan gözlemleyemeyiz ancak o özelliği ölçtüğünü varsaydığımız testlerle söz konusu özelliği doğrudan değil, dolaylı olarak ölçebiliriz. Bu durumda da yapılan ölçme sonucunda elde edilen sayı o özelliğin gözlenen değerini oluşturur. Bu sayıya da gözlenen puan denir (Baykul, 2010).

Eğitim ve psikolojideki ölçümlerde, bireyin davranışları gözlemlendiğinden ölçme sonuçlarına mutlaka belirli oranlarda hata karışır. Hata, ölçülmek istenmeyen değişkenlerin ölçme sonuçlarına karışmasıdır. Eğitim ve psikolojide bireyin gerçek puanlarına bir miktar karışan hata değeri, gerçek puana eklenerek bireyin gözlenen puanı elde edilmiş olur. Eşitlik (1)'de belirtilen durumun matematiksel ifadesine yer verilmiştir (Cohen ve Swerdlik, 2005).

$$X_{\text{gözlenen}} = X_{\text{gerçek}} + X_{\text{hata}} \quad (\text{Crocker ve Algina, 2008}). \quad (1)$$

Ayrıca KTK'de varyans gerçek puan ile hatanın kareleridir ve güvenilirlik katsayısı gerçek puanın sahip olduğu varyansın gözlenen puanın sahip olduğu varyansa oranı olduğu eşitlik (2)'de verilmiştir.

$$\begin{aligned} \sigma_{\text{gözlenen}}^2 &= \sigma_{\text{gerçek}}^2 + \sigma_{\text{hata}}^2 \\ \rho &= \sigma_{\text{gerçek}}^2 / \sigma_{\text{gözlenen}}^2 \end{aligned} \quad (\text{Crocker ve Algina, 2008}). \quad (2)$$

Hatayı düşük bir seviyeye indirebilmek hata kaynaklarının belirlenmesi ile kontrol edilebilmektedir. Hatalar; ölçme amacıyla kullanılacak test, sınav ya da ölçekten, ölçmeye konu olan nesne ya da ölçülecek bireylerden, ölçme sürecini uygulayan ölçmeci ya da ölçmecilerden, ölçmenin yapıldığı ortam koşullarından ve ölçme aracının puanlanmasından ve bu puanların yorumlanmasından kaynaklanmaktadır (Cohen ve Swerdlik, 2005).

Ölçme işlemlerinde bulunan değişkenler, hata varyanslarına sebep olabilmektedir. Hatalar, sabit, sistematik ve tesadüfi olarak üç türde incelenebilir (Erkuş, 2003). Bir

ölçmeden diğerine, ölçme sonucuna etki ettiği oran değişmeyen hatalar sabit hatalardır. Sabit hatalar, her bir ölçmede aynı miktar ve taraftadır (Baykul ve Turgut, 2011; Erkuş, 2003). Örneğin bir öğretmenin bir testteki bütün cevap kağıdına fazladan 5 puan vermesi ölçmeler için hatalı ama hata yönü ve miktarı aynı olacaktır.

Sistemik hatalar belirli bir yönü iyileştirici ya da tam tersi miktarlarda görülebilirler. Ölçülen grubun büyüklüğüne, ölçme işlemine veya ölçmenin uygulandığı koşullara bağlı olarak sistemik bir şekilde değişen hata türlerine denir (Baykul ve Turgut, 2011; Erkuş, 2003). Açıklamayı örnekleyecek olursak aynı öğretmenin yazı güzelliğinin ölçülmediği bir testte yazısı güzel olanlara fazladan 5 puan vermesi, diğerlerine ise vermemesi sistemik hatayı ortaya çıkarır.

Ölçme sonuçlarına rastgele etki eden, nereden geldiği, ne kadar büyüklükte olduğu ve etkisinin ne tarafta etkili olduğu kesin bir şekilde bilinmeyen hatalar tesadüfi hatalardır (Baykul ve Turgut, 2011). Testi alan ya da puanlayıcıdan kaynaklanan dikkatsizlik, yorgunluk, sağlık durumu, motivasyon, psikolojik durumu gibi hata kaynakları olabileceği gibi ortamın ısısının, sesli olmasının ya da ışığın testi alan kişileri etkilemesi de tesadüfi hata kaynakları olabilir. Bu tür hata kaynaklarının ne olduğu kesin bir şekilde bilinemez ve büyüklüğü hesaplanamaz ancak kestirilebilir. Sabit ve sistemik hata kaynakları önceden önlenemez fakat psikolojideki hataların miktarı, yönü ve kaynağı çoğu zaman belli değildir ve bu nedenle önlenemez (Erkuş, 2003).

Ölçme çalışmalarında, gözlenen puanların gerçek puanlara yakın olarak elde edilebilmesi amaçlanmaktadır. Gözlenen puanların hatalardan arınlık derecesinin yüksek olduğu durumlarda gerçekte olması beklenen sonuçlara yakın ölçme sonuçları ortaya çıkmaktadır. Tesadüfi hataların az olması ise ölçmenin güvenilir olup olmadığı ile bilgi verir. Güvenirliğin kestirilebilmesi için geliştirilmiş Klasik Test Kuramı (KTK), Madde Tepki Kuramı (MTK) ve Genellenabilirlik Kuramı (GK) kuramları vardır. MTK’de madde parametreleri gruptan bağımsız olduğu için testlerin güvenilirlikleri hata kaynağı olan birey göz önünde bulundurulmadan kestirilebilmektedir fakat MTK’nin yerine getirilmesi güç varsayımlarının olması kuramın kullanımının sınırlı yönlerindedir (Hambleton ve Swaminathan, 1985). Bu kuramlara dayalı güvenilirlik kestirme yöntemleri ve bu yöntemler üzerine uygulanabilecek istatistiksel hesaplamalar vardır.

Klasik Test Kuramında Güvenirlik ve Güvenirliğin Kestirilme Yöntemleri

Cohen ve Swerdlik (2005) güvenirliği, “ölçmelerin tutarlılığı” olarak

tanımlamaktadır. Erkuş (2003) güvenilirliğini, ölçme araçlarının olabilecek en az seviyede hata ile ölçme yapabilmesi olarak tanımlamaktadır. Turgut (1990), “ölçme sonuçlarının tesadüfi hatalardan arınlık derecesi” olarak da tanımlamaktadır. Anastasi (1982), güvenilirliği aynı ölçme aracını farklı durum ve zamanlarda alan aynı bireylerin ölçme sonuçlarının tutarlılığı olarak tanımlamıştır.

KTK’ye göre güvenilirlik, güvenilirlik katsayısı kavramı ile matematiksel olarak ifade edilebilir. Güvenirlik katsayısı, ölçmelerdeki hatalardan arınlığı belirtir ve bir testin paralel iki kısmından elde edilen puanların arasındaki korelasyon olarak belirtilmektedir. KTK’ye göre güvenilirlik kestirme yöntemleri vardır. Güvenirlik kestirim yöntemleri de ölçmelerde elde edilecek sonuçların hangi amaçla kullanılacağına ve ölçmelerdeki sonuca etki eden değişkenlerden hangisinin hata olarak belirlenmesi amacıyla kullanılmaktadır (Amerikan Psikoloji Birliği, 2014). Aynı ölçme aracının, aynı bireylere farklı zamanlarda aynı koşullar altında uygulanmasından elde edilen korelasyon katsayısını belirleyen test tekrar test yöntemidir. Belirlenen bu katsayı aynı zamanda kararlılık katsayısı olarak da bilinir. Test tekrar test yönteminde korelasyon katsayısını düşüren zaman problemi vardır. Testin ikinci kere uygulanması arasındaki zaman kısa tutulduğunda testi alan bireylerin test maddelerini ve cevaplarını hatırlama olasılığı olduğu için iki uygulama arasındaki korelasyon katsayısı düşer. Testin ikinci kere uygulanması arasındaki zaman uzun tutulduğunda ise ölçülen özelliği unutma ya da ölçülen özellik ile ilgili özel ders alma gibi değişkenler sonuçları etkileyebilir. Dolayısıyla aynı testin uygulanması arasındaki zaman arttıkça, testlerden elde edilen puanlar arasındaki ilişki düşer. Bu nedenle test-tekrar test güvenilirliği zaman içerisinde kişilik gibi daha az değişime uğrayacak özelliklerin değerlendirilmesinde kullanılması uygun olan bir güvenilirlik tahminidir. (Aiken, 2000; Anastasi, 1982; Cohen ve Swerdlik, 2005).

Ölçülmek istenen özelliği ölçen farklı maddelerden oluşan iki eşdeğer (paralel) formun farklı zamanlarda aynı bireylere uygulanmasıyla uygulanan bir yöntem eşdeğer formlar yöntemidir. Bir teste eşdeğer form denilebilmesi için her iki testin de içeriklerinin, madde sayılarının ve tiplerinin aritmetik ortalamalarının, standart sapmalarının eşit olması gerekmektedir. Eşdeğer formlar güvenilirliği, eşdeğerlik katsayısı hesaplanarak kestirilmektedir. Bu katsayının yeteri kadar yüksek oluşu her iki formun da güvenilirliğinin yüksek olduğuna, yeteri kadar yüksek olmayışı da her iki formun da güvenilirliğinin düşük olduğuna delil olarak sayılmaktadır. Eşdeğer formlar güvenilirliğine üç durumda hata kaynakları karışabilmektedir: (1) aynı gruba iki test uygulamak gerekmektedir, (2) test puanları motivasyon, öğrenme, terapi gibi faktörlerden

etkilenebilmektedir, (3) madde örnekleme testini alan bireylerin gerçek yeteneklerinin bir fonksiyonu olmadan bireylerin lehine ya da aleyhine çalışmasıdır (Aiken, 2000; Baykul, 2010; Cohen ve Swerdlik, 2005).

Test, bir grup bireye sadece bir kez uygulanarak, testin kendi kendisiyle tutarlılığına bakılır. Bu bakımdan güvenilirlik değerlendirme işlemleri iç tutarlılık, bulunan güvenilirlik katsayısı da iç tutarlılık katsayısı olarak isimlendirilir. Tek bir testin iki eşdeğer yarılarından elde edilen iki puanın birbirleriyle ilişkilendirilmesiyle eşdeğer yarılar güvenilirlik kestirimi elde edilmiş olur. İki testin kullanılmasının pratik olmadığı ya da istenmediği durumlarda uygulanan bir yöntemdir. Eşdeğer yarılar yöntemi testin içeriğinin örneklemeyle ilişkili bir tutarlılık ölçüsü sağlamaktadır. Bu nedenle bu yöntem iç tutarlılık katsayısının hesaplanması olarak da adlandırılır. Eşdeğer yarılar güvenilirlik katsayısı üç aşamada hesaplanmaktadır.

1. Test eşdeğer yarıya bölünür,
2. Testin iki yarısından elde edilen iki puan üzerinde Pearson Momentler Çarpımı Korelasyon Katsayısı “r” hesaplanır,
3. Eş değer yarılar güvenirligi Spearman Brown formülü kullanarak hesaplanır.

Yarı yarıya bölmenin temel amacı, her bir yarının diğerine eşit olduğu mini paralel formlar oluşturmaktır ama bu formlar oluşturulurken yeterince paralellik oluşturulamaması hata kaynağıdır (Aiken, 2000; Anastasi, 1982; Cohen ve Swerdlik, 2005).

Eşdeğer yarılar inde kullanılan Spearman Brown formülü, yarıya bölünen iki test arasında elde edilen korelasyondan iç tutarlılık katsayısının hesaplanmasına olanak sağlamaktadır. Bir testin güvenirligi, testin uzunluğundan etkilendiği için kısaltılmış ya da uzatılmış testin güvenirligini kestirmek için bu formüle ihtiyaç vardır. Testin uzunluğu iki katına çıkarılırsa sahip olacağı güvenirligi yansıtacak şekilde düzeltilmeli ve güvenirligi yukarı doğru ayarlanmalıdır. Bu düzeltme için kullanılan formüle Spearman Brown düzeltme formülü denir ve eşitlik (3)’te bu formül gösterilmiştir:

$$r_w = (2 \cdot r_h) / (1 + r_h) \quad (3)$$

“ r_w ”, Spearman Brown formülü tarafından ayarlanan tüm testin güvenirligine, “ r_h ” ise iki yarıya bölünmüş testin puanları arasındaki Pearson Momentler Çarpımı Korelasyon Katsayısı “r” ye eşit olmaktadır (Kubiszyn ve Borich, 2007).

Eşdeğer yarılar yönteminde güvenilirliğin kestirilmesinde kullanılan formüller Kuder-Richardson-20 ve Kuder-Richardson-21 formülleridir. Bu formüller, maddelerin (1-0) olarak ikili puanlanmasına dayanan testlere uygulanmaktadır ve testin tamamına ilişkin güvenilirlik katsayısı hesaplanmasında kullanılmaktadır. Maddelerin testin bütünüyle olan iç tutarlılığının bir ölçüsüdür. Testteki her maddenin aynı değişkeni ölçtüğü dolayısıyla testin homojen olduğu sayılımasına bağlıdır.

Cronbach Alfa yöntemi de testin iç tutarlılığının hesaplanabildiği bir yöntemdir. KR 20-21 yöntemine benzerdir ancak çoklu puanlanabilen maddelerden oluşan testlere uygulanabilmektedir. Test tekrar test güvenilirlik kestirimini yapabilmek için uygulanabilmektedir (Cronbach, 1990).

Güvenirlik, ölçmelerin hatalardan arınlık derecesi olduğuna göre ölçmeye karışan hatanın ölçüsü de ölçmenin standart hatasıdır. Bu ölçüm, bireysel puanların yorumlanmasında bazı durumlarda güvenilirlik katsayısına göre daha kullanışlı olmaktadır. Güvenirlik katsayısı bir korelasyon katsayısı olarak bire doğru yaklaştıkça hata oranı o ölçüde azalmakta sifira yaklaştıkça hata oranı o ölçüde artmaktadır. Hata oranı da güvenilirlik katsayısına göre artıp azalmaktadır. Ölçmenin standart hatasını hesaplayabilmek için güvenilirlik katsayısını bilmek gerekmektedir.

$$\text{Ö}_{sh} = S\sqrt{(1 - r)} \quad (4)$$

Eşitlik (4)'te ifade edilen S test puanlarının standart sapması, r ise güvenilirlik katsayısını ifade eder (Anastasi, 1982).

KTK aynı anda tek bir hata kaynağı ile ilgilenebildiği için güvenilirlik kestirim yöntemleri de sadece bir hata türüne göre farklılık içermektedir. Test-tekrar test yöntemi uygulanarak elde edilen güvenilirlik söz konusu olduğunda zaman içerisinde değişim önemli olmaktadır ve hata kaynağı zaman olmaktadır. Paralel testler yöntemi ile elde edilen güvenilirlik ölçme aracının benzer testleri arasındaki tutarlılık önemlidir ve buradaki hata kaynağı da testlerdir. İç tutarlılık ile ilgili güvenilirlik hesaplama yöntemlerinde ise ölçme aracında bulunan maddelerin birbiri ile tutarlılıkları önemlidir ve bu yöntemde hata kaynağı maddelerdir (Anastasi, 1982; Kline, 2000; Cohen ve Swerdlik, 2005). KTK'de güvenilirlik kestirim yöntemlerinin bahsedilen hata kaynaklarından ötürü sınırlılıkları vardır.

Klasik Test Kuramında Kestirilen Güvenirliklerin Sınırlılıkları

Bir test kuramı, ölçme sürecinde problemlerin ortaya koyulması ve bu problemlere uygun çözümlerin veya açıklamaların getirilmesi ya da azaltılmasına yönelik çalışmaları içerir. KTK'nin de eğitim ve psikoloji de en yaygın kullanılan kuram olmasına karşın bazı sınırlılıkları vardır. Bu sınırlılıklar aşağıda sıralanmıştır.

- a) Klasik test kuramında, madde istatistikleri gruba bağımlıdır. Bir teste yönelik güvenilirlik ve geçerlik değerleri, uygulandığı gruba bağlı olarak farklı değerler almaktadır.
- b) Klasik test kuramında, madde alanı örnekleme teorisi esas alındığından, puanların güvenilirliğini kanıtlamak için aynı madde evreninden gelen paralel bir teste ihtiyaç duyulmaktadır. Ancak paralel testlerin oluşturulması çok zor olmaktadır. Çünkü eşit olduklarını kanıtlamak için büyük çalışmalar gerektirmektedir. (Gulliksen,1950). Bu paralel formlar, sadece ortalama değil, standart sapmalar, güvenilirlik, ayırt edicilik ve geçerlilik ölçütlerinde de benzer olmalıdırlar.
- c) Klasik test kuramıyla geliştirilen testler, genellikle orta güçlükte maddelerden oluşmaktadır ve orta seviyede yetenek grubuna hitap etmektedir ve yüksek veya düşük yetenek seviyesindeki bireyleri belirlemede yetersiz kalmaktadır. Bu yüzden büyük gruplarda ve seçme amaçlı kullanılmasında sınırlı kalmaktadır.
- d) Güvenirlik, grup puanları için bile ayrı ayrı elde edilemez. Bireysel puanlar için ayrı güvenilirlik kestirimi yoktur. Bu da bireyleri farklı farklı etkileyen durumlar sonucu oluşan ve farklı bireylere ait olan farklı hata miktarının yok sayılıp, hesaplanan tek bir hata değerinin grup üzerine genellendiğini göstermektedir.
- e) Klasik test kuramı, ölçmeye karışan tek hata olarak tesadüfi hata ile ilgilenir, sistematik hata ve sabit hatayı göz ardı etmektedir (Hambleton ve Swaminathan, 1985).

Klasik test kuramında bulunan güvenilirlik kestirimleri ve hata varyanslarının puanlar üzerindeki etkilerinin neler olduğu ve bu etkilerin hangi oranda olduğunun belirlenmesinde yetersiz kaldığı görülmüştür. Bu konu üzerinde çalışanların farklı hata varyanslarını nasıl belirleyeceği üzerine yeni fikirler geliştirmelerine yol açmıştır. Aynı zamanda KTK'nin örneklem grubuna bağlı olarak istatistikler ortaya koyması sonuçların evrene uyarlanamaması gibi sınırlılıkları G Kuramına dayanak olmuştur (Crocker ve Algina, 1986; Eason, 1989). G kuramı, KTK'deki ayrıştırılmamış hatanın potansiyel

ölçüm hatası kaynaklarının büyük oranda açıklanmasını sağladığı kabulünden ortaya çıkmıştır (Shavelson ve Webb, 2005).

Genellenebilirlik Kuramı

Genellenebilirlik kuramı (GK), gözlemlenen puanlarda ortaya çıkan tek bir hata kaynağı yerine ölçmeye etki eden birçok hata varyansının incelenmesine olanak sağlayan kapsamlı bir kavramsal çerçeve ve güçlü istatistiksel hesaplamalar içeren bir kuramdır. Bir araştırmacının, ortaya çıkan veya ortaya çıkacak hata varyanslarını ve gözlemlenen puanlardaki tutarsızlık kaynaklarını tespit edip ölçebilmesine olanak sağlar (Brennan, 2001). G kuramını Cronbach, Rajaratman, Nanda ve Gleser (1963 ve 1972) gibi araştırmacılar tarafından geliştirilmiş ve Crocker ve Algina (1986), Algina (1989), Brennan ve Kane (1979), Feldt ve Brennan (1989), Suen (1990), Shavelson, Webb ve Rowley (1989), Shavelson ve Webb (1991), Kieffer (1999), Marcoulides (1999,2000) Strube (2000) ve Brennan (2001)'in katkılarıyla geliştirilmiştir (Mushquash and O'Connor, 2006; Li, Shavelson, Yin ve Wiley, 2015).

Gözlenen puanlarda var olan hatalar farklı kaynaklardan ortaya çıkabilir. Bu farklı hata kaynaklarını aynı anda ve tek bir katsayı ile ortaya çıkarma ve karşılaştırma yolu G kuramıdır. G kuramı, varyans oluşturan bütün ölçüm hatalarını ayrıştırır, tahmin eder ve ortaya çıkarır ve bulunan bu hatalar varyans bileşenleri olarak adlandırılır (Mushquash and O'Connor, 2006). G kuramı, her bir kaynağın katkıda bulunduğu varyans tahminlerini sağlayarak çeşitli kaynaklar arasındaki etkileşimlerle ilişkili varyans tahminlerini belirler (Brennan, 2001).

G kuramı, araştırmacının aşağıdaki soruları ele almasına izin verir.

1. Standart ölçme hatasını belirlemek için madde ve puanlayıcı örnekleme yeterli midir?
2. Madde ya da puanlayıcı sayısını artırarak ölçümün güvenilirliğini artırabilir miyim?
3. Test puanları bireyin başarısına ilişkin yargıda bulunmak konusunda güvenilir midir? (Shavelson ve Webb, 2005).

Yukarıdaki soruların cevapları ölçme sürecinde kestirilmeye çalışılan güvenilirlik katsayısının önemli bir testte bireyler hakkında kararlar alınırken daha güvenilir değerlendirmeler yapılmasına olanak sağlamaktadır. Aynı zamanda psikoterapi gibi birçok doğrudan gözlemlenemeyen değişkenlerin etki ettiği alanlarda süreç

değerlendirme ölçütlerinin güvenilirliğini değerlendirmek çok önemlidir. G kuramı, bahsedildiği gibi birden fazla ölçüm hatası kaynağı eş zamanlı olarak değerlendirilebilir (Wasserman, Levy ve Loken, 2009).

G kuramı, KTK'yi özel bir durum olarak ele alır ve KTK'nin kavramlarını ve ölçüm tasarımlarını barındırır (Eason, 1989). Bu nedenle G kuramı ile ilgili bir kuramsal çerçeve sunarken KTK'nin hangi eksikliklerine ya da farklılıklarına yönelik olarak G kuramının geliştirildiğinin ifade edilmesi gerekmektedir.

G Kuramı ve KTK'nin Sınırlı ve Güçlü Yönlerinin Karşılaştırılması

G kuramı, geleneksel güvenilirlik kavramlarının daha geniş bir çerçevede kullanılmasına olanak verip, güvenilirliği varyans bileşenlerine ve tahminlerine dayandırarak varyans analizine (ANOVA) imkan sağlamaktadır. KTK'nin, aynı anda tek bir hata kaynağı ile ilgili çıkarımlar yapabilmesine dayanan bir kuram olmasının sınırlılığını gidermek için geliştirilmiş genellebilirlik kuramı, birçok hata kaynağının aynı anda incelenip tek bir güvenilirlik kestiriminin yapılabilmesine olanak sağlamaktadır. G kuramı bireyin gözlenen puanına KTK'den farklı olarak birçok sayıda hata karıştığını ifade eder. Eşitlik (5)'te de belirtildiği gibi gözlenen puan, evren puanına ve sonsuz sayıda hata kaynağının toplanmasıyla elde edilir.

$$\begin{array}{c} \text{Hata kaynakları} \\ \text{-----} \\ X = \mu_P + E_1 + E_2 + \dots + E_N \quad (5) \\ \swarrow \quad \searrow \\ \text{Gözlenen puan} \quad \text{Evren puanı} \end{array}$$

Çok sayıda hata kaynağı, G kuramında temel bir kavram olan genelleme evrenini tanımlar (Shavelson, Webb ve Rowley, 1988). KTK'nin temelini oluşturan güvenilirlik kavramı G kuramında yerini, daha geniş ve esnek genelleme kavramına bırakır. KTK'de ifade edilen gerçek puan G kuramında evren puanıdır. Hata ise tek bir hata kaynağının etkisi varyans bileşenleri olarak ifade edilmektedir. Bu ikisinin toplamı da G kuramına göre gözlenen puanı vermektedir (Eason, 1989).

G kuramının KTK'nin genişletilmiş bir uzantısı olduğunu Shavelson ve Webb (1991), VanLeeuwen (1997), Brennan (2001) ve Wang (2005) beş maddeyle

belirtmişlerdir.

1. Genellenebilirlik kuramı tek bir analizle birçok hata kaynağını kestirebilmektedir.
2. Değişkenlik kaynaklarının her birinin büyüklüğünü belirleyebilmektedir.
3. Bireylerin elde ettiği başarılarına yönelik mutlak ve bağıl değerlendirmeler aynı anda alınabilmektedir ve buna bağılı olarak iki farklı güvenilirlik katsayısı hesaplanabilmektedir.
4. İstenilen ölçme durumlarında, en uygun güvenilirlik katsayısının elde edilebileceği karar çalışmaları yapılabilmektedir.
5. Gelecekte yapılacak uygulamalarda etkin ölçüm prosedürleri tasarlanabilmektedir.

KTK, güvenilirliği değerlendirmek için yararlı yöntemler (test-tekrar testi, paralel formlar, iç tutarlılık hesaplamaları vb.) sunsa da bir seferde ancak bir ölçüm hatası kaynağı olarak belirlenebilmektedir. Ayrıca birçok güvenilirlik tahmini elde etmek için mümkün olan çok sayıda güvenilirlik katsayısını birleştirmenin, çeşitli hata kaynaklarının göreceli bir önemini belirlemenin ya da hata kaynakları arasındaki olası etkileşimleri değerlendirebilmenin bir yolu yoktur. Araştırmacılara, katılımcılardan güvenilir puanlar elde etmek için kaç tane madde, test ve gözlemci gerektiğine dair açık bir bilgi vermez (Webb, Rowley, Shavelson, 1988). G kuramı, KTK'nin uzantısı olarak kabul edilse de bazı yönlerden araştırmacılara ölçme sonuçlarını ya da ölçme araçlarını geniş kapsamlı değerlendirebilmesi imkanı verdiği için çeşitli farklılıkları vardır. Aşağıda G kuramının, KTK'ye göre üstün yönleri verilmiştir.

1. G kuramı, ölçme sürecinde yer alan hata kaynaklarının hepsini aynı anda ele alınıp yorumlanabilmesine olanak sağladığından, aynı anda sadece bir adet hata kaynağının yorumlanabildiği KTK'ye göre daha geniş yorumlanabilen bir güvenilirlik kestirimi yapılabilmesini mümkün kılar.
2. G Kuramı, KTK'deki gibi sadece ölçmedeki hata varyanslarının değişkenlik etkilerine değil, bu hata varyans değişkenliklerinin ortak etkilerini de göz önünde bulundurur.
3. G kuramı, birçok hata kaynağının birleşimini aynı zamanda test-tekrar test ve iç tutarlılık güvenilirliği ve puanlayıcılar arası güvenilirliklerin hesaplanıp birlikte değerlendirilmesine olanak sağlamaktadır.
4. G kuramında, ölçmenin güvenilirliğinin kestirilmesinde bağılı ve mutlak değerlendirmeler için hesaplamalar yapılabilirken KTK'de ise sadece

bağlı değerlendirme için güvenilirlik kestirilebilir.

5. G kuramında karar çalışmaları ile istenilen her değişkenlik kaynağının koşullarının sayısının değiştirildiği durumda güvenilirliğin sınanmasına imkan tanır. KTK'de ise sadece bir değişkenlik kaynağının (madde/görev) sayısının değiştirilmesinin güvenilirliği nasıl etkilediği Spearman-Brown formülü ile hesaplanıp elde edilebilir (Shavelson & Webb, 1991; VanLeeuwen, 1997; Brennan, 2001).

G kuramı, genellebilirlik çalışması (G çalışması) ile Karar (Desicion) çalışması (K ya da D çalışması) olmak üzere iki ayrı analiz aşamasını içerir. İlk aşama, sonuçların bir popülasyon için genelleştirilebilir ve genelleştirilebilirlik çalışmasını (G çalışması) analiz eder. İkinci aşama ise G çalışmasından elde edilen bilgileri kullanan karar çalışmasını (K çalışması) içerir.

Kabul Edilebilir Gözlemler ve G Çalışması

Bir değişkeni ölçmek için yola çıkan bir araştırmacı, çoğu zaman gözlemini ilgili alan ya da evrene genelleme eğilimindedir (Cronbach, 1990). Bireyin gerçek puanlarının tahminleri ve kabul edilebilir gözlemleri bir evrenin parçalarıdır. G çalışması ise, ölçümlerin evrene genellenmesiyle ilgilenmektedir. Evren ise ölçüm örnekleminin daha geniş bir ölçüm koşulu olarak tanımlanmaktadır. Ölçüm sonuçlarının istikrarı veya bir ölçekteki maddelerle ilgili çalışmaların hepsi G çalışmaları olarak kabul edilebilir. G çalışmasının temel amacı, değişkenlik kaynakları hakkında elde edilebilecek bütün bilgileri elde etmeye çalışarak hata varyansının büyüklüğünü hesaplamaya çalışmaktır. G çalışması için öncelikle ölçme örnekleminin genelleneceği evren ve değişkenlik kaynakları tespit edilmelidir. G çalışması varyans analizine dayanır. Varyans oluşturan her bir etmene değişkenlik kaynağı (facet, yüzey) denir ve G çalışması olabildiğince fazla değişkenlik kaynağı içerecek şekilde oluşturulmalıdır.

G kuramına göre bireysel gözlem ve ölçme, yalnızca olası ölçümlerin sonsuz ve çeşitli evrenlerinden elde edilen bir tahminidir. Bireyin gerçek puanlarının tahminleri ve kabul edilebilir gözlemler evreninin parçalarıdır (Brennan, 2001). Test formları, maddeler, ölçme durumları ve puanlayıcılar potansiyel bir hata kaynağı olan ölçme yönteminin herhangi bir özelliğini içerebilen (facet) yüzey olarak tanımlanabilir. Her yüzeyin içindeki farklı maddeler ya da ölçme durumları gibi düzeyler sonsuz derecede büyük olabilen koşullardır. Ölçümün amacı genellikle bir hata kaynağı olarak kabul

edilmeyen bireylerdir çünkü bireyler değişkendir ve bireyler arasında gerçek puan farklılıklarının olması beklenmektedir. Ölçme objesi bir hata sapması yaratmadığı için yüzey olarak kabul edilmez (Mushquash and O'Connor, 2006). Evren puanı, olası seçenekler evrenindeki bir birey üzerindeki tüm gözlemlerin beklenen bir değeri olarak tanımlanır, bu ise KTK'deki gerçek puana eşdeğerdir. Evren veya genelleme evreni, ilgilenilen yapıyı değiştirmeden bir ölçümün kabul edilebilir bütün yüzeylerini, paralel olarak kabul edilebilecek her yüzey için seviyeleri belirlenerek tanımlanır.

Örneğin, bir araştırmacı bir grup test maddesini daha geniş bir grup test maddesine genellese, değişkenlik kaynağı testteki maddeler olmaktadır. Eğer araştırmacı bir test formunu daha çok sayıda bir test formuna genellemek istiyorsa bu örnekteki değişkenlik kaynağı ise test formları olur. Bütün kabul edilebilir test formları ise evren olarak tanımlanır.

Sonsuz Genelleme Evreni ve Karar (K) Çalışması

K çalışmaları, G çalışması ile belirlenerek yorumlanan varyans bileşenlerinin, kullanımı ve yorumlanması hakkında karar verilebilmesi için bilgi elde edebilmek amacıyla uygulanır. K çalışması, en güvenilir ölçüm süreçlerini elde etmek ve G çalışmasında uygulanabilecek en uygun ölçme desenine karar verebilmek için G çalışmasında elde edilen bilgileri kullanır (Brown, 2005). K çalışmasının uygulanmasındaki amaç, güvenilir sonuçlar elde edebilmek için değişkenlik kaynaklarının sayısı belirlenip hatanın en aza indirilmesidir.

Araştırmacı ölçüm yöntemlerinin sonuçlarına dayanarak genelleme evreni tanımlar. Genelleme evreni, genellemek istediği koşullar setidir. Evren puanı, genelleme evrenindeki ölçümlerin ortalamasıdır. Evren puanının, beklenen gözlem puanı varyansına oranı genellenebilirliğin etkili olduğunun bir ölçüsüdür. G kuramı, bir değerlendirmenin belirli kararlar için uyarlanabileceğini kabul eder ve G çalışmasını K çalışmasından ayırır. G çalışmasında evren kabul edilebilir gözlemlerin değerlendiricilere varyans bileşen tahminleri sağlamak için mümkün olduğunca geniş bir şekilde tanımlanırken K çalışması ise belirlenmiş bir hedef için sadece hedefle ilgili olan yüzeyleri belirler, böylece puanların yorumlanması genelleme evrenine yönelik yapılabilir. Genellenebilirlik katsayısı, her bir yüzey için ayrıca hesaplanabilir (Shavelson ve Webb, 2005).

G kuramında, araştırmacının amacına uygun olarak yüzey ve koşullarına ve değişkenlik kaynaklarının sayısına göre uygun desen seçilmelidir. Seçilen desenleri

uygulayarak genellenebilirlik katsayısı hesaplanır. G çalışması için G ve Phi katsayısı hesaplanır (Brennan, 2001; Crocker ve Algina, 1986).

G kuramında arařtırmacılar, G ve K çalışmalarını uygularken seçtikleri desenler vardır. Bu desenler, arařtırmanın hedefine, arařtırmada kullanılacak verilere, arařtırmadaki deęişkenlik kaynaklarının türüne, arařtırmanın sonuçlarına ve nihayetinde kestirilmiş olan güvenirlilik katsayısına göre seçilen farklı desen ve yöntemler vardır (Shavelson ve Webb, 2005; Brennan, 2001; Shavelson ve dięerleri, 2015).

Çaprazlanmış ve Yuvalanmış Desenler

Bir deęişkenin tüm yüzeyleri dięer deęişkenlik kaynağının tüm yüzeylerinde görüldüğünde elde edilen tasarım çaprazlanmış desendir. Çapraz tasarlanmış bir desende bir yüzeyin tüm koşulları dięer her yüzeyin tüm koşullarında gözlenir. Örneğin tek yüzeyli tasarlanmış bir desende her bir birey, her bir madde üzerinde ölçülür ve b x m olarak gösterilir (Shavelson ve Webb, 2005; Huebner ve Lucht, 2019).

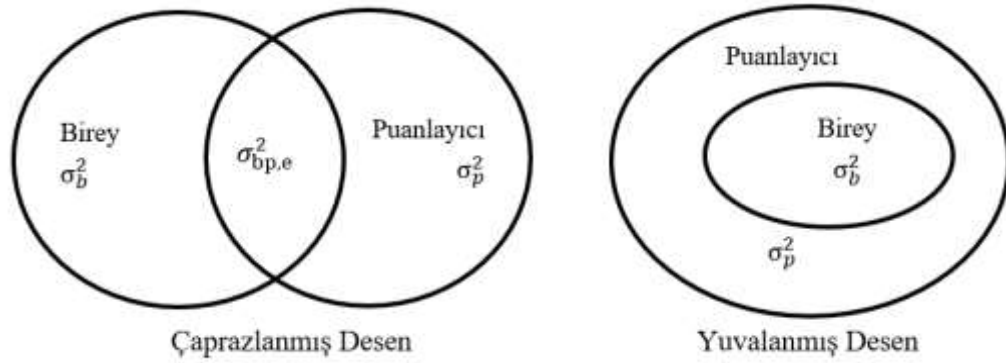
Bir deęişkenlik kaynağının bazı yüzeyleri dięer deęişkenlik kaynağının tüm yüzeylerinde gözlemlenmiyorsa yuvalanmış desendir. Yuvalanmış olarak tasarlanmış bir desende bir yüzeyin bütün koşulları dięer her yüzeyin bütün koşullarında gözlemlenmez. Örneğin tek yüzeyli tasarlanmış bir desende bütün bireyler, bütün maddeler üzerinde ölçülmediyse yuvalanmış desen olur ve b : m olarak gösterilir (Shavelson ve Webb, 2005; Huebner ve Lucht, 2019).

Örneğin, bir arařtırmada bir ölçek birden fazla puanlayıcı tarafından puanlanması sonucunda ařağıda belirtilen dört farklı durum ortaya çıkar ve G ve K çalışması için uygun desenlerden biri kullanılabilir (Crocker ve Algina, 1986)

1. Her birey bir puanlayıcı tarafından puanlanır ve bu puanlayıcı her bireye puan verir.
2. Her birey birden fazla puanlayıcı tarafından puanlanır ve her puanlayıcı her bireye puan verir.
3. Her birey birbirinden farklı puanlayıcılar tarafından puanlanır ve tek bir bireyi yalnızca bir puanlayıcı deęerlendirir.
4. Her birey birden fazla puanlayıcı tarafından puanlanır ve her birey için farklı puanlayıcılar vardır.

Yukarıdaki dört desende de tek deęişkenlik kaynağı puanlayıcıdır. İlk iki desende, bütün öğrenciler aynı puanlayıcılar tarafından puanlandığı için ölçmenin deęişkenlik

kaynağı olan puanlayıcı birey ile çaprazlanmıştır. Çaprazlanmış desen puanlayıcı p ile ve birey b ile gösterilmek suretiyle pxb şeklinde gösterilir. Çaprazlanmış bir desenden elde edilen veriye eşleştirilmiş veri (matched data) denir. Son iki desende ise her birey farklı puanlayıcılar tarafından puanlandığından puanlayıcılar bireylerle yuvalanmıştır. Yuvalanmış desen ise puanlayıcı p ile ve birey b ile gösterilmek suretiyle b:p şeklinde gösterilir. Yuvalanmış desenden elde edilen veriye bağımsız veri (independent data) denir. Birbirinden farklı bireyler (B), başka maddeleri (M) cevaplandığı ve birçok farklı puanlayıcının (P) farklı öğrenciler tarafından cevaplanan farklı maddeleri puanlaması ile oluşan tasarıma ise tümüyle yuvalanmış desen adı verilir. Bu desende bireyler puanlayıcılarla, maddeler de bireylerle yuvalanmıştır ve (B : M : P) olarak ifade edilir (Cronbach, Gleser, Nanda, & Rajaratnam, 1972; Crocker ve Algina, 1986). Bazı ölçme durumlarında da değişkenlerin çaprazlanmış ve yuvalanmış olarak belirlendiği desenler de uygulanabilmektedir (Brennan, 2001; Shavelson ve diğerleri, 2015). Çaprazlanmış desende, etkileşim ve ana etki bileşenlerinin ayrı ayrı tahminlerine olanak verirken, yuvalanmış desende bazı varyans bileşenlerinin ayrı tahminlerine olanak vermez (VanLeeuwen, 1997). Şekil 1’de çaprazlanmış ve yuvalanmış desene ait Venn şeması gösterilmiştir. Bu şema ile çaprazlanmış ve yuvalanmış desenin değişkenlik kaynakları, bu değişkenlik kaynaklarının birbiriyle etkileşimleri ifade edilmektedir (Brennan, 2001).



Şekil 1. Tek Yüzeyle Çaprazlanmış ve Yuvalanmış Desen Örneği Venn Şeması (Brennan, 2001).

Bu çalışmada ölçmenin hedefi bireyler olacağı için bireyler değişkenlik kaynağı ya da yüzey olarak adlandırılmaz. Bireyler çalışmada ölçmenin objesi olarak adlandırılır. Değişkenlik kaynakları madde ve puanlayıcılardır. Madde ve puanlayıcıların seviyeleri de koşullar olarak adlandırılmaktadır. Araştırmada kullanılan değişkenlere ait koşullar,

arařtırmacının amacına ve ilgilendiđi kısımlara bađlı olarak sabit ya da tesadüfi olabilir. Sabit deđiřkenin tüm kořulları evrende sabittir bu durumda arařtırmacı sonuçlarını evrene genelleyemez. Arařtırmacı tesadüfi deđiřkenlik kořullarını ise olası gözlemler evreninden tesadüfi olarak seçilip örneklendirildiđi ve puanlayıcı evreninden tesadüfi puanlayıcı seçilmesi ve bunların puanlama yapması ile örneklemden daha büyük bir evrene genelleme yapabilir (Crocker ve Algina, 1986; Brennan, 2001).

G kuramında, en az bir facet olası deđiřkenlik oluřturacak bir kaynak olarak belirlenmelidir. Belirlenecek facet, ölçmeden elde edilen sonuçların genellenmesi planlanan kısımdır. Buna “tek deđiřkenlik kaynaklı evrenler” (tek yüzeyle evren) denir (Yelbođa ve Tavřancıl, 2010). Tek deđiřkenlik kaynaklı evrenlerde ölçmeye karıřan hata tek bir hata kaynađından gelmektedir. Tek deđiřkenlik kaynaklı evrenler üzerine kurulan desenlerde dört tane varyans kaynađı bulunmaktadır. Birincisi bireylerin bilgi, beceri, tutum gibi benzeri farklılıklarından kaynaklanmaktadır. Bu varyans kaynađı ölçmenin amacı olarak ifade edilmektedir. Deđiřkenlik kaynađından gelen tek bir hata kaynađı olan evren, bir deđiřkenlik kaynaklı evren olarak tanımlanmaktadır. Varyansın ikinci kaynađı ise ölçme aracından kaynaklanmaktadır. Ölçme aracında bulunan maddelerin kolay, zor ya da orta güçlükte olması varyans kaynađı olarak belirtilmektedir. Varyansın üçüncü kaynađı ise, bireylerin geçmiř deneyimleridir. Bireylerin ilgisini çeken konularla ilgili maddeleri cevaplama olasılıkları daha yüksektir. Varyansın dördüncü kaynađı ise tanımlanamayan veya kaynađı bilinmeyen deđiřkenler olarak adlandırılmaktadır. Birinci ve ikinci varyans kaynakları olan birey (b) – madde (m) ve ikisinin etkileřimi (bxm) ile tesadüfi veya tanımlanamayan varyans kaynakları (e) birleřir ve birlikte artık olarak kabul edilir (Crocker ve Algina, 1986; Brennan, 2001; Shavelson ve diđerleri, 2015).

Bazı durumlarda ölçüm kořulları iki veya daha fazla deđiřkenlik kaynađı içerebilmektedir. İki deđiřkenlik kaynađının belirlendiđi durumlar iki yüzeyle evrenler olarak tanımlanmaktadır. Bireylerin, farklı maddelerden oluřan bir testi birden fazla puanlayıcının puanlaması ile oluřturulan çalıřma, iki deđiřkenlik kaynaklı bir çalıřma örneđidir. Puanlayıcılar ve test maddelerinin her biri bir deđiřkenlik kaynađıdır (Crocker ve Algina, 1986). İki deđiřkenlik kaynaklı evrenlerde varyans bileřenleri ařađıdaki gibi ifade edilmiřtir.

Tablo 1

İki yüzeyli evrenlerde varyans bileşenleri

Değişkenlik Kaynağı	Değişkenlik Türü	Varyans Sembolü
Birey (b)	Bireyin varyansı	σ^2_b
Madde (m)	Maddelerin varyansı	σ^2_m
Puanlayıcı (p)	Puanlayıcıların varyansı	σ^2_p
Birey x Madde	Birey ve madde etkileşiminin varyansı	σ^2_{bm}
Birey x Puanlayıcı	Birey ve puanlayıcı etkileşiminin varyansı	σ^2_{bp}
Madde x Puanlayıcı	Madde ve puanlayıcı etkileşiminin varyansı	σ^2_{mp}
Birey x Madde x Puanlayıcı, e	Artık varyans	$\sigma^2_{bmp, e}$

(Crocker ve Algina, 1986; Brennan, 2001; Güler, Kaya Uyanık ve Taşdelen Teker, 2012)

İki değişkenlik kaynaklı evrenlere bir örnek verilecek olursa;

Bir araştırmada kabul edilebilir gözlemler evreni P (birey), T (test), R (puanlayıcı) olmak üzere P x T x R belirlenmiş olsun. Bu durum için tek bir puanlayıcı tarafından değerlendirilen tek bir test için gözlemlenebilir herhangi bir puan aşağıdaki eşitlikte temsil edilmektedir:

$$X_{PTR} = \mu + v_p + v_T + v_R + V_{PT} + V_{PR} + V_{TR} + V_{PTR} \quad (6)$$

Eşitlik (6)'ya göre bu tasarım için belirlenen ortalama puanı ve birbiriyle ilişkisiz yedi bileşeni içermektedir. Eşitlik (6)'nın verdiği ortalama puanları ve kabul edilebilir gözlemler evrenindeki koşullar üzerinde fark varyans etkilerinin belirtildiği eşitlik (7)'deki denklemde verilmiştir ve gözlemlenen toplam puan farkı yedi bağımsız varyans bileşenine ayrıştırılabilir.

$$\sigma^2(X_{ptr}) = \sigma^2(p) + \sigma^2(t) + \sigma^2(r) + \sigma^2(pt) + \sigma^2(pr) + \sigma^2(tr) + \sigma^2(p) \quad (7)$$

Eşitlik (7) incelendiğinde varyans bileşenlerine rastgele varyans bileşenleri demek doğru olacaktır. Tablo 2'deki desen incelendiğinde iki yüzeyli bir desen olduğu

görülmektedir. Bu desen göz önüne alındığında eşitlik (7)'deki varyans bileşenlerinin tahmini için Tablo 2'deki Beklenen Ortalama Kareler denklemleri kullanılmaktadır. Bu bilgi araştırmacıya en büyük değişkenliği sağlayan kaynakları tanımlamasını sağlar (Huebner ve Lucht, 2019).

Tablo 2

Beklenen ortalama kareler ve P x T x R için tahmin edilen varyans bileşenleri

Etki (α)	Beklenen Ortalama Kareler (α)
P	$\sigma^2(\text{ptr}) + n_t \sigma^2(\text{pr}) + n_r \sigma^2(\text{pt}) + n_t n_r \sigma^2(\text{p})$
T	$\sigma^2(\text{ptr}) + n_p \sigma^2(\text{tr}) + n_r \sigma^2(\text{pt}) + n_p n_r \sigma^2(\text{t})$
R	$\sigma^2(\text{ptr}) + n_t \sigma^2(\text{tr}) + n_t \sigma^2(\text{pr}) + n_p n_t \sigma^2(\text{r})$
Pt	$\sigma^2(\text{ptr}) + n_r \sigma^2(\text{pt})$
Pr	$\sigma^2(\text{ptr}) + n_t \sigma^2(\text{pr})$
Tr	$\sigma^2(\text{ptr}) + n_p \sigma^2(\text{tr})$
Ptr	$\sigma^2(\text{ptr})$
Etki (α)	$\sigma^2(\alpha)$
P	$[\text{MS}(\text{p}) - \text{MS}(\text{pt}) - \text{MS}(\text{pr}) + \text{MS}(\text{ptr})] / n_t n_r$
T	$[\text{MS}(\text{t}) - \text{MS}(\text{pt}) - \text{MS}(\text{tr}) + \text{MS}(\text{ptr})] / n_p n_r$
R	$[\text{MS}(\text{r}) - \text{MS}(\text{pr}) - \text{MS}(\text{tr}) + \text{MS}(\text{ptr})] / n_p n_t$
Pt	$[\text{MS}(\text{pt}) - \text{MS}(\text{ptr})] / n_r$
Pr	$[\text{MS}(\text{pr}) - \text{MS}(\text{ptr})] / n_t$
Tr	$[\text{MS}(\text{tr}) - \text{MS}(\text{ptr})] / n_p$
Ptr	MS (ptr)

Sosyal bilimlerdeki ölçme işlemleri ölçmeye konu olan özellikler bakımından daha karmaşıktır ve her zaman iki değişkenlik kaynağı ile ele alınmayabilir. İki den fazla değişkenlik kaynağının araştırmaya konu olduğu durumda araştırma, üç ve daha çok yüzeyli değişkenlik kaynaklı evrenler olarak tanımlanabilir. Örneğin araştırmacı bir test sonucunu kullanacaksa maddeleri, zamanı ve test uygulayıcısı gibi üç değişkenlik kaynağına genellemek isteyebilir (Shavelson ve Webb, 1991).

Değişkenlik kaynağı arttıkça, örneklemden elde edilecek verilerle evrene yönelik yorumlanması planlanan genellemede de hata yapma olasılığı artar. Dolayısıyla

araştırmaya etki edecek hata varyanslarının bilinmesi gerekmektedir (Güler, Kaya Uyanık ve Taşdelen Teker, 2012).

Hata Varyansları

Hata varyansları sonsuz genelleme evreni göz önüne alındığında ölçme objesi dışındaki varyans bileşenleri bir veya daha fazla farklı hata varyansı türüne katkıda bulunur. Bunlar mutlak (absolute) ve bağıl (göreceli, relative) varyans türleridir.

Parriott (2016)'a göre KTK'ye göre G kuramını kullanmanın farklılıklarından biri de G kuramıyla iki hata varyansı hesaplanabilmesidir. G kuramına göre bağıl hata ve mutlak hata varyansı bulunmaktadır. Mutlak hata varyansı sadece bir kişinin gözlenen ve evren puanı arasındaki farktır ve " $\sigma^2(\Delta)$ " şeklinde gösterilir. Gruptan bağımsız olarak mutlak bir değere göre bireylerdeki değişiklikler kestirilir. Örneğin; (bxm_{xp}) desenli bir çalışmada mutlak hata varyansı aşağıdaki formül kullanılarak hesaplanır (Brennan, 2001; Shavelson ve Webb, 1991)

$$\sigma^2(\Delta) = \frac{\sigma_m^2}{n_m} + \frac{\sigma_p^2}{n_p} + \frac{\sigma_{bm}^2}{n_m} + \frac{\sigma_{bp}^2}{n_p} + \frac{\sigma_{mp}^2}{n_m n_p} + \frac{\sigma_{bmp}^2}{n_m n_p} \quad (8)$$

Bağıl (göreceli) hata, bireyin gözlenen ve evren sapma puanı arasındaki farktan elde edilmektedir. Gözlenen ve evren puanları uygulandığı örneklemin ortalamalarına bağıl olarak hesaplandığı için göreceli (relative) ismini alır (Güler, Kaya Uyanık ve Taşdelen Teker, 2012). Bağıl hata varyansı göreceli değerlendirmeler yapılacağı zaman ve bireylerin grup içerisindeki durumlarını birbiriyle karşılaştırarak değerlendirmek istendiğinde kullanılmaktadır. " $\sigma^2(\delta)$ " şeklinde gösterilir (Brennan, 2001). Örneğin; b x m x p desenli bir çalışmada bağıl hata varyansı aşağıdaki formül kullanılarak hesaplanır (Brennan, 2001; Shavelson ve Webb, 1991).

$$\sigma^2(\delta) = \frac{\sigma_{bm}^2}{n_m} + \frac{\sigma_{bp}^2}{n_p} + \frac{\sigma_{bmp}^2}{n_m n_p} \quad (9)$$

Genellenebilirlik Katsayısı ve Güvenirlik Katsayısı

G kuramında hangi modelin uygun olacağı ölçme sonuçlarının araştırmacı

tarafından nasıl kullanılacağına bağlı olarak değişir. G kuramı, birçok G katsayısı ile ilgilenir ve evrenin tanımlanmasının değişmesiyle katsayı değerlerinin de değişeceğini belirtir. Evren tanımı bireyin grubun içinde bulunduğu durumdan bağımsız ya da bireyin grubun içinde bulunduğu durum önem arz ederek ifade ediliyorsa farklı iki güvenilirlik katsayısı hesaplanabilmektedir. Gruba bağlı yapılan ölçümler için genellenebilirlik katsayısı (G), gruptan bağımsız yapılan ölçümler için ise güvenilirlik katsayısı (Phi) hesaplanmaktadır. Eğer bir araştırmacı bireyi değerlendirirken bireyin performansını grup içindeki başarısına göre değerlendiriyorsa genellenebilirlik katsayısını (G) göz önünde bulundurmalıdır.

Genellenebilirlik katsayısının uygun bir şekilde formüle edilebilmesi genelleme evrenine ve karar çalışmasının tasarlandığı şekle bağlıdır (Crocker ve Algina, 1986). Genellenebilirlik katsayısı “ E_{p^2} ” şeklinde gösterilir.

$$E_{p^2} = \frac{\sigma^2(b)}{\sigma^2(b) + \sigma^2(\delta)} \quad (10)$$

$\sigma^2(b)$: Evren Puanı Varyansı

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

Eşitlik (10) incelendiğinde, payın evren puanı varyansına ($\sigma^2(b)$) eşit, paydanın da $\sigma^2(b)$ ile $\sigma^2(\delta)$ değerlerinin varyanslarının toplamına eşit olduğu görülmektedir. Genellenebilirlik katsayısı 0 ile 1 arasında değer alır. Eğer G katsayısı düşük bulunursa madde veya puanlayıcı sayıları artırılmalıdır (Güler, Kaya Uyanık ve Taşdelen Teker, 2012).

Eğer araştırmacı bireyleri değerlendirirken mutlak bir ölçüte bağlı olarak değerlendirme yapıyorsa bireylerin performansı gruptan bağımsız olarak değerlendirilir ve güvenilirlik katsayısını (Phi) kullanır. Güvenirlik katsayısı (Phi), genellenebilirlik katsayısına (G) göre daha katıdır ve hem ölçüm nesnelerinin sıralama düzenindeki tutarlılık derecesini hem de ham puanların yüksekliklerindeki tutarlılığını yansıtır. Phi katsayıları, elde edilen puanların gerçek değerleri araştırmacı için anlamlı olduğunda yararlıdır ve kabul edilen bir kesme değerinin olduğu performans ölçümlerinde (direksiyon testleri veya psikiyatrik tanımlar için) anlamlıdır.

Güvenirlik katsayısı (Phi) aşağıdaki formül ile hesaplanabilir:

$$\Phi = \frac{\sigma^2(b)}{\sigma^2(b) + \sigma^2(\Delta)} \quad (11)$$

$\sigma^2(b)$: Evren Puanı Varyansı

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

Eşitlik (11) incelendiğinde, payın evren puanı varyansına ($\sigma^2(b)$) eşit, paydanın da $\sigma^2(b)$ ile $\sigma^2(\Delta)$ değerlerinin varyanslarının toplamına eşit olduğu görülmektedir. Araştırmada kullanılan madde ve puanlayıcı sayısının artması Phi katsayısının değerini de artıracaktır. Mutlak hata varyansı, bağıl hata varyansına göre daha çok varyans bileşeni içerdiğinden görece hata varyansına göre büyük değerler alacaktır dolayısıyla Phi katsayısı G katsayısından her zaman daha küçük değere sahip olacaktır (Alharby, 2006).

Göreceli kararlar için G tahminlerine yalnızca kişilerin sıralama düzenini yansıtan varyans bileşenleri dahil edilirken, mutlak kararlar için G tahminlerine tüm varyans bileşenler (evren puanları hariç) dahil edilir (Musquash ve O'Connor, 2006).

G kuramında, hata varyanslarından biri de maddelerdir. Bir ölçekte, maddelerin açık uçlu maddeler olarak hazırlanması güvenilirlik hesaplamalarında bir hata kaynağı dolayısıyla değişken olarak değerlendirilmelidir. Aynı zamanda açık uçlu maddelerin değerlendirilmesinde puanlayıcıların verdiği puanların ve bu puanlara dayanarak personel alımı sınavında bireyler hakkında verilen kararların güvenilirliğini belirlemek gerekmektedir. Bu sebeple açık uçlu maddelerin güçlü ve sınırlı yönlerinin incelenmesi bu madde türleri üzerinde uygulanacak G kuramının anlaşılmasını kolaylaştıracaktır.

Açık Uçlu Maddeler

Açık uçlu maddeler cevap üretmeye dayalı sorulardır. Bireylere açık uçlu maddeler sorularak onların bilgilerini organize etme, sentezleme, değerlendirme ve birleştirme potansiyellerini ortaya çıkarmaları beklenir (Thorndike, 2005). Açık uçlu maddelerle bireylerin öğrendikleri bilgileri yeni durumlarda kullanarak yaratıcı fikirler geliştirmeleri beklenmektedir.

Açık uçlu hazırlanmış maddelerin diğer madde türlerine göre bazı üstün yönleri vardır. Popham (1990)'a göre açık uçlu maddeler üst düzey öğrenme ürünlerini ölçmek için kullanılan tartışmasız en iyi yaklaşımdır. Çünkü bu tip maddeler öğrencilerden düşünceleri orijinal bir şekilde bir araya getirmeleri ister. Öğrencilerin birbirlerinden

farklı bu cevapları, seçmeli cevaplı ya da kısa cevaplı maddelerde elde edilemez. Açık uçlu maddelerden oluşan sınavlarda cevaplar öğrenci tarafından organize edilerek yazılması gerektiği için doğru cevabın tahmin edildiği sınavlara göre daha zordur. Açık uçlu maddeler öğrencilerin kompozisyon yazma yeteneğini geliştirmelerini sağladığı için çok yaygın kullanılır. Her açık uçlu maddede öğrenciler kompozisyon yazma yetenekleri ile ilgili pratik yapmış olurlar. Bu nedenle açık uçlu maddelerin yazma yeteneği üstünde olumlu etkileri vardır. Açık uçlu maddelerin diğer bir üstün yanı, oluşturulmalarının kolay ve az zaman alıcı olmasıdır. Çünkü açık uçlu maddelerin asıl zorluğu değerlendirme kısmındadır

Açık uçlu maddelerin ölçme aracı olarak birçok sınırlılığı vardır. Başarılı bir ölçme gerçekleştirebilmek için ölçme araçlarının güvenilirlik, geçerlik ve kullanılabilirlik olmaz üzere üç önemli özelliği vardır ve ölçme araçları bu özellikleri bakımından sürekli karşılaştırılırlar (Hopkins, 1998). Açık uçlu maddelerin sınırlılıkları halo etkisi, sorudan soruya taşıma etkisi, testten teste taşıma etkisi, sıralama etkisi ve dil ve düzen etkisi kısımları altında sıralanmıştır (Popham, 1990).

Açık uçlu maddelerin sınırlılıklarından kaynaklı birçok hata varyansı vardır ve bu değişkenler güvenilirliği etkilemektedir. Yukarıdaki hata kaynaklarının yanı sıra önemli bir hata kaynağı da puanlayıcılarıdır. Açık uçlu maddeler, puanlayıcılara subjektif puanlama yapmaya olanak sağladığı için birden fazla puanlayıcının olduğu durumlarda puanlayıcılar arası güvenilirliğin belirlenmesinin gerekli olduğu pek çok ölçme durumu vardır. KTK' de sadece tek bir hata kavramıyla ilgilenilebildiği için G kuramının puanlayıcı, birey, ölçme formu, maddeler gibi birçok hata varyansının aynı anda hesaplanabilmesine olanak vermesi açık uçlu soruların sınırlılıklarını gidermekte uygun bir yöntem olduğu belirtilmiştir. Bu amaçla personel alımı sınavında uygulanmış açık uçlu maddelerden oluşmuş ve maddelerin iki puanlayıcı tarafından değerlendirildiği bir ölçekte birey, puanlayıcı ve madde değişkenlik kaynaklarının hepsinin aynı anda G kuramı uygulanarak analiz edilip yorumlandığı çalışmaya gereksinim duyulmuştur.

Alanyazın incelendiğinde, G kuramına dayalı farklı desenlerin olduğu araştırmalar (Nalbantoğlu, 2009; Nalbantoğlu Yılmaz, 2012; Alkan, 2013; Doğan ve Anadol, (2016); Taştan, 2017) bulunmaktadır. Nalbantoğlu (2009), araştırmasında çaprazlanmış ve yuvalanmış deseni karşılaştırarak G ve K çalışmalarını yorumlamıştır. Nalbantoğlu Yılmaz (2012), araştırmasında yuvalanmış tasarlanmış bir desenin dengelenmiş ve dengelenmemiş desenlere göre G çalışması sonuçlarını karşılaştırmıştır. Alkan (2013), araştırmasında açık uçlu maddelerden elde edilen veri üzerine iki farklı desen

uygulayarak G ve K çalışmasını yapmıştır. Doğan ve Anadol (2016), araştırmalarında çaprazlanmış ve yuvalanmış deseni karşılaştırarak G ve Phi katsayılarını yorumlamışlardır. Taştan (2017), araştırmasında üç farklı çok yüzeyli deseni karşılaştırarak G çalışması yapmıştır. Buna göre iki yüzeyli çaprazlanmış ve yuvalanmış desenin birlikte uygulandığı iki ayrı desenin G ve K çalışmaları ile belirlenen G ve Phi katsayı değerlerinin doğru yorumlanabilmesinin literatüre katkı getireceği düşünülmektedir. Bu nedenle bu araştırmanın problemi, açık uçlu maddelerden oluşmuş personel alımı sınavına katılan bireylerin cevaplarının, farklı puanlayıcılar tarafından değerlendirilmesiyle elde edilen verinin genellenebilirlik kuramındaki farklı desenler uygulanarak genellenebilirlik (G) ve karar (K) çalışmalarının yapılması ve sonuçlarının analiz edilip yorumlanmasıdır.

Amaç

Bu araştırmanın amacı, Ankara Üniversitesi Sınav Yönetim Merkezi'nin uyguladığı açık uçlu maddelerden oluşmuş personel alımı sınavına katılan bireylerin cevapladıkları maddelerin, farklı puanlayıcılar tarafından puanlanmasının genellenebilirlik kuramındaki desenlerden çaprazlanmış ve yuvalanmış desenin birlikte uygulanmasına göre yapılacak G ve K çalışmalarının yapılması ve bu çalışmaların sonucunda elde edilecek G ve Phi katsayılarının analiz edilip yorumlanmasıdır. Bu amacı gerçekleştirebilmek için aşağıdaki sorulara cevap aranmıştır.

Alt Amaçlar

1. Sınava katılan birey (b), sınavdaki maddeler (m) ve maddeleri değerlendiren puanlayıcı (p) değişkenlerinin tümüyle çaprazlanmış olarak tasarlandığı (bxm_{xp}) deseninin genellenebilirlik kuramı sonuçları nasıl kestirilmiştir?
 - 1.1. (bxm_{xp}) deseninin G çalışması ile belirlenen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıl kestirilmiştir?
 - 1.2. (bxm_{xp}) deseninde madde ve puanlayıcı sayılarının artırılması ve azaltılmasıyla yapılan karar (K) çalışması sonuçları nasıl kestirilmiştir?
2. Birey (b) değişkeninin çapraz, puanlayıcı (p) ve madde (m) değişkenlerinin yuvalanmış tasarlandığı (bx(m:p)) deseninin genellenebilirlik kuramı sonuçları nasıl kestirilmiştir?

- 2.1. (bx(m:p)) deseninin G çalışması ile belirlenen varyans bileşenleri ve toplam varyansı açıklama yüzdeleri nasıl kestirilmiştir?
- 2.2. (bx(m:p)) deseninde madde ve puanlayıcı sayılarının artırılması ve azaltılmasıyla yapılan karar (K) çalışması sonuçları nasıl kestirilmiştir?
3. (bxmxp) ve (bx(m:p)) desenlerinin uygulanması ile kestirilen G çalışması sonuçlarının karşılaştırması nasıldır?
4. (bxmxp) ve (bx(m:p)) desenlerinin uygulanması ile kestirilen karar (K) çalışması sonuçlarının karşılaştırması nasıldır?

Önem

Açık uçlu maddelerden oluşmuş üst düzey düşünme becerilerinin ölçüldüğü sınavlar, öğretimin her aşamasında önem taşımaktadır. Özellikle bu araştırmanın konusu olan bir meslek alanına ilişkin bilgilerin ölçüldüğü personel alımı sınavında bireylerin edindikleri bilgilerin ne düzeyde olduğunu belirlemek ve bu doğrultuda bireyler hakkında alınacak kararların güvenilirliği çok önemlidir.

Bu çalışmada da ASYM tarafından yapılan açık uçlu maddelerden oluşan personel alımı sınavının, uygulanması sonucunda bireylerin cevaplarının iki puanlayıcı tarafından puanlanmasıyla elde edilen ölçümlerin güvenilirliklerini genellenebilirlik kuramına göre belirlemek ve bu kurama göre uygun desenler oluşturmak önemlidir. Ayrıca G kuramından elde edilen G ve Phi güvenilirlik katsayılarının doğru yorumlanması önemlidir. Karar (K) çalışması sonucu ölçümlerin güvenilirliğinin iyileştirilmesine yönelik önerilerin de fayda sağlayacağı düşünülmektedir.

Alanyazın incelendiğinde Cronbach (1990)'ın G kuramında oluşturulacak desenlerde bütün varyans kaynaklarına ilişkin hata tahminleri yapımına izin verdiğinden çaprazlanmış desenin uygulanmasını önermektedir. Ancak Nalbantoğlu (2009), çaprazlanmış ve yuvalanmış desenlerin G ve Phi katsayılarının karşılaştırıldığı tez çalışmasında yuvalanmış desende bu katsayıların daha büyük çıkma eğiliminde olduğunu bulmuştur. Ayrıca Alkan (2013)'ün tez çalışmasında tümüyle çaprazlanmış ve yuvalanmış deseni karşılaştırarak ortaya koyduğu üzere G ve Phi katsayılarının yuvalanmış desende daha büyük çıktığı aynı zamanda varyans bileşenlerinin hesaplanmasında serbestlik derecesini artırdığı belirtilmiştir. O nedenle bu çalışmada oluşturulan yuvalanmış desenin G ve Phi katsayılarını yorumlamak için yeni bir bakış açısı oluşturacağı ve bu desende de yüksek güvenilirlik katsayıları elde edilebileceği

düşünülmektedir. Aynı zamanda çok sayıda öğrencinin bulunduğu açık uçlu maddelerden oluşmuş sınavlarda puanlayıcılar arası tutarlılık sağlandığında puanlayıcıların bireylerin hepsini puanlaması yerine belirli sayıdaki bireylerin sırayla puanlanmasının işgücü ve zaman bakımından ekonomik olacağı sonucu desteklenmiş olacaktır. Bu şekilde ileri seviye personel alımı sınavında belirli aşamalardan geçmiş bireylerin oluşturduğu örnekleme, madde havuzundan seçimli cevaplanacak sınavlarda farklı maddelerin analizlerinden de güvenilir sonuçlar elde edilebileceği düşünülmektedir.

Sayıtlar

İşe giriş sınavına giren bireylerin, maddeleri yanıtlarken sınava gerçek performanslarını yansıttıkları varsayılmıştır. Puanlayıcıların, puanlama yaptıkları süreçte dikkatli davrandıkları varsayılmıştır.

Sınırlılıklar

Araştırma, ASYM'nin yaptığı personel alımı sınavına katılan 211 katılımcı ile sınırlıdır. Araştırmada, G kuramı birey, madde ve puanlayıcı değişkenlik kaynakları kullanılarak uygulanmıştır. G kuramının uygulanmasında kullanılacak diğer değişkenlik kaynakları (test formları, cinsiyet, zaman vb.) araştırma süreçlerine katılmamıştır. Araştırmada, açık uçlu maddelerden oluşan bir sınav kullanıldığı için pratiklik açısından puanlayıcılar iki ile sınırlandırılmıştır.

İlgili Araştırmalar

Bu başlık altında G kuramındaki farklı desenlerin incelendiği ve farklı desenler üzerinde yapılan çalışmaların karşılaştırıldığı araştırmalara yer verilmiştir.

Yurtiçinde Yapılan Araştırmalar

Atılgan (2004), Müzik Öğretmenliği Özel Yetenek Seçme Sınavları verileri üzerinde G kuramı ve Rasch Modeli ile kestirilen istatistiki sonuçlarını karşılaştırmıştır. Bu yaklaşımlar ile elde edilen sonuçların her durumda paralellik göstermediği, durumdan duruma farklılaşabildiği sonucu elde edilmiştir. Ancak özellikle puanlayıcı ve görev

değişkenlik kaynağının her iki yaklaşım içinde tutarlı sonuçlar ürettiği görülmüştür. Atılğan (2004)'in çalışmasına benzer şekilde Güler (2008) de başka bir veri üzerinde KTK, G kuramı ve Rasch modelini uygulayarak çalışmasında kullanmıştır.

Güler (2008), KTK, G kuramı ve Rasch Modeli uygulayarak yaptığı araştırmada matematik başarısının ölçülmesiyle sonuçların güvenilirliklerini hesaplamış ve üç farklı yöntemden ortaya çıkan sonuçları karşılaştırmıştır. G kuramını için ise, bxgxp tümüyle çaprazlanmış desen kullanılarak G ve Phi katsayıları kestirilmiştir. Sonuçlara göre, araştırmadaki ölçme aracından güvenilir değerler elde edildiği, puanlayıcıların birbirleriyle tutarlı oldukları bulunmuştur. Ayrıca puan güvenilirliği kestirilmesinde, puanların hangi amaçla kullanılacağına göre üç kuramdan içinden uygun olanın tespit edilmesi ve güvenilirlik hesaplama yöntemine göre en az iki kuramdan yararlanıldığında sonuçların güvenilir olacağı sonucuna ulaşmıştır. Güler (2008), çalışmasında KTK, G kuramı ve Rasch modelinin karşılaştırılması üzerinde durduğu; Deliceoğlu (2009)'da çalışmasında ise KTK ve G kuramını karşılaştırmıştır.

Deliceoğlu (2009), araştırmada performans ölçen bir futbol yetilerine ilişkin dereceleme ölçeğinin KTK ve G kuramına göre kestirilen güvenilirlik katsayılarının belirleyip karşılaştırmıştır. G kuramını uygulamak için madde ve puanlayıcı sayılarını artırıp azaltmış ve G ve Phi katsayılarını hesaplamıştır. Buna göre G katsayısının beklenen değerinden yüksek, Phi katsayısının ise beklenen değerinden düşük olduğu görülmüştür. Hata kaynaklarının fazla olduğu bu araştırmada G kuramının, KTK'ye alternatif oluşturduğu yorumu yapılmıştır. Yelboğa ve Tavşancıl (2010)'da araştırmalarında KTK ve G kuramını karşılaştırmışlardır.

Yelboğa ve Tavşancıl (2010), araştırmalarında bir iş performansı ölçeğinden elde edilen verilerle KTK ve G kuramı analiz etmiştir. KTK ve G kuramına göre kestirilen güvenilirlik katsayılarını karşılaştırarak yorumlamışlardır. G kuramını uygulamak amacıyla çok değişkenlik model uygulanarak G ve Phi katsayıları hesaplanmıştır. Bulgulara göre aynı ölçme durumu için KTK'nin ve G kuramının çok değişkenlik model uygulanarak elde edilen güvenilirlik katsayılarının birbirleriyle uyumlu olduğu sonucuna ulaşmışlardır. Güler'de (2011), Yelboğa ve Tavşancıl (2010) gibi G kuramı ve KTK'nin rastgele veriler üzerinde güvenilirliğin karşılaştırıldığı bir araştırma yürütmüştür.

Güler (2011), araştırmada G kuramı ve KTK'ye göre güvenilirliği rastgele veriler üzerinde karşılaştırmıştır. Tek değişkenlik kaynaklı çaprazlanmış tasarlanan desen (b x m) için elde edilen G katsayısı ile küçük katsayı değerlerine ulaşmıştır. Güvenilir olmayan veri üzerinde çalışılarak elde edilen sonuçların beklendiği gibi çıkması hem G

kuramına hem de KTK'ye göre desteklenmiş olduğu yorumunda bulunmuştur. Büyükkıdık'da (2012), Güler (2011) gibi matematiksel problemi çözme becerilerinin değerlendirilmesine yönelik iki farklı puanlama anahtarına göre G kuramı ve KTK'yi kullanarak puanlayıcılar arası güvenilirliği hesaplamıştır.

Büyükkıdık (2012), iki adet performans görevinin, analitik ve bütünsel dereceli puanlama anahtarıyla, dört puanlayıcıdan elde edilen sonuçlara KTK ve G kuramı uygulayarak puanlayıcılar arası güvenilirliği karşılaştırmıştır. Sonuçlara göre G kuramının sonuçlarına göre G ve Phi katsayılarının KTK'deki katsayılara göre bağıl olarak büyük çıkma eğiliminde olduğu, G kuramının uygulanmasının KTK'ye göre daha detaylı bilgi sağladığı ve her iki kuramdan da elde edilen sonuçların verilerin güvenilirliğinin yüksek olduğunu desteklediği görülmüştür. Başat (2015), Büyükkıdık (2012)'ın araştırması ile paralel bir şekilde öğrencilerin performanslarının ölçülmesinde kullanılan üç farklı ölçme aracının güvenilirliğini G kuramına göre değerlendirmiştir.

Başat (2015), G kuramına göre öğrencilerin performanslarının ölçülmesinde uygulanan üç farklı ölçme aracının G ve D çalışmaları sonuçlarını karşılaştırmıştır. Öğrencilere, araştırmacı tarafından geliştirilmiş çoktan seçmeli test, yapılandırılmış grid ve performans görevi uygulanarak yapılandırılmış grid ve performans görevi değerlendirilmesinde üç puanlayıcı görev almıştır. K çalışmaları sonucunda her üç ölçme aracı için elde edilen G ve Phi katsayılarının değişiminin benzerlik gösterdiği belirlenmiştir ve yapılandırılmış grid ve performans görevini değerlendiren üç puanlayıcı arasındaki korelasyona bakılarak puanlayıcıların öğrencileri puanlamada aralarında farklılıklar olmadığı bulunmuştur.

Bağcı (2015), araştırmasında ortaokul yedi sınıfı öğrencilerine matematiksel muhakeme performansının tespit edilmesinde uygulanan ölçekten elde edilen ölçümlerin güvenilirliğinin KTK ve G kuramının çaprazlanmış ve yuvalanmış desenleri uygulayarak karşılaştırılmıştır. Araştırmanın çalışma grubuna, matematiksel muhakeme seviyelerini belirleyen ölçek uygulanarak öğrenci cevapları üç bağımsız puanlayıcı tarafından analitik puanlama anahtarı ile puanlanmıştır. Araştırma sonucuna göre her iki kuramdan kestirilen güvenilirlik katsayıları da kabul edilebilir düzeyde bulunmuştur. (öxs:p) deseninde G çalışması sonucu kestirilen G ve Phi katsayıları (öxsxp) deseninden daha yüksek bulunmuştur. Her 3 puanlayıcı için ayrı ayrı hesaplanarak elde edilen Cronbach alfa katsayıları ise, her iki desende bağıl ölçme için kestirilen G katsayıları ile oldukça paraleldir. Ayrıca Genellenebilirlik kuramında yapılan karar çalışması ile de yüzeylerin sayılarının mutlak ve bağıl hata varyanslarına etkisi belirlenmiştir. Dolayısıyla G kuramı

ile yapılan analizlerin KTK'ye göre yapılan analizlerden daha detaylı bilgi verdiği görülmüştür. Bağcı (2015) gibi Şalgam (2016)'da araştırmasında G kuramı ve KTK kuramını karşılaştırmıştır.

Şalgam (2016), 7. Sınıf öğrencilerine yönelik olarak hazırlanmış “Kısa Cevaplı Matematik Yazılı Sınavı”ndan elde edilen ölçümlerin güvenilirliğinin Klasik Test Kuramının Test Tekrar Test Yöntemi ve Genellenebilirlik Kuramının çaprazlanmış deseniyle kıyaslanmıştır. Ulaşılan güvenilirlik katsayılarının arasında anlamlı bir farklılık bulunmamıştır. Pekin, Çetin ve Güler (2018)'de Şalgam (2016) gibi KTK ve G kuramı sonuçlarını karşılaştırmıştır.

Pekin, Çetin ve Güler (2018), araştırmalarında Otizim Sosyal Beceriler Profili ölçeğinin beş puanlayıcı tarafından değerlendirilmesiyle ulaşılan puanları KTK ve G kuramına göre karşılaştırarak puanlayıcılar arası güvenilirliği yorumlamışlardır. G kuramı için çaprazlanmış ve yuvalanmış desen uygulanarak G ve Phi katsayıları hesaplanmıştır. Sonuçlara göre G kuramının KTK'ye göre daha ayrıntılı sonuçlara ulaşıldığı, her iki kuramında Otizim Sosyal Beceriler Profili ölçeğinin otizimli bireylerin düzeylerini tespit etmede güvenilir sonuçlara ulaşıldığı aynı zamanda her iki kuram için hesaplanan puanlayıcılar arası tutarlılığın düşük olarak elde edildiği belirlenmiştir.

Nalbantoğlu Yılmaz ve Gelbal (2011), G kuramında uygulanan farklı desenleri karşılaştırdıkları araştırmalarında öğrencilerin puanlayıcılar tarafından bir arada ve dönüşümlü olarak puanlanmasıyla oluşturulan tümüyle çaprazlanmış ve bireylerin puanlayıcılara yuvalanarak oluşturulan desenlerin G ve K çalışmalarının sonuçlarını karşılaştırmışlardır. Her iki desene kestirilen değerlerinin benzer olduğu, K çalışmaları sonucunda yuvalanmış desende G ve Phi katsayılarının daha büyük elde edilme eğilimde olduğu gözlemlenmiş ve sonuçta puanlayıcıların belirli sayıda öğrencilerin sırayla değerlendirilmesinin iş gücü ve zaman açısından daha ekonomik olacağı sonucuna varmışlardır. Alkan (2013)'te Nalbantoğlu Yılmaz ve Gelbal (2011) gibi G kuramıyla farklı desenleri uygulayıp karşılaştırmıştır.

Alkan (2013), PISA 2009 Okuma Becerileri performansını belirlenmesi için uygulanan açık uçlu maddelerden oluşan testin puanlayıcıların hepsinin ve bazı öğrencileri bölüşerek puanlamasıyla elde edilen desenlere göre G kuramı karşılaştırmasını yapmıştır. ÖxSxP çaprazlanmış deseni ve Ö:PxS çaprazlanmış ve yuvalanmış deseni birlikte uygulanmıştır. Çaprazlanmış ve yuvalanmış desenin birlikte uygulandığı desen ile elde edilen bağıl ve mutlak hata varyanslarının çaprazlanmış desene göre daha düşük elde edildiği dolayısıyla G ve Phi katsayı değerleri daha büyük çıkmıştır.

Desenler üzerinde uygulanan K çalışmaları yorumlandığında puanlayıcı sayısı artırmanın G ve Phi katsayılarında artışa neden olduğu görülmüştür. Anadol (2017)'de, Nalbantoğlu Yılmaz ve Gelbal (2011) ve Alkan (2013) gibi G kuramında farklı desenleri uygulayıp G çalışması sonuçlarını karşılaştırmıştır.

Anadol (2017), araştırmasında İngilizce yazma becerisinin değerlendirilebilmesi amacıyla, rubrik kullanılması üzerine üç grup puanlayıcıdan toplanan verilerin G çalışmaları sonuçlarını karşılaştırmıştır. Araştırma, rubrik kullanmaya ilişkin deneyimi az, deneyimi çok ve rubrik kullanmaya ilişkin deneyimi az ve çok olan puanlayıcıların aynı grupta yer alması ile oluşan karışık grup olmak üzere üç ayrı grupta aynı dereceli puanlama anahtarı ile 6 görev doğrultusunda, (b : p) x m deseni ile öğrencileri değerlendirdiği G çalışmaları yapılmıştır. Araştırma sonuçlarına göre üç farklı durumda elde edilen değişkenler için kestirilen varyans değerleri, G ve Phi katsayıları, mutlak ve bağıl hata varyansları bütün gruplarda benzer elde edilmiştir. Bu durumda, kalifiye bir rubrik kullanarak yapılan güvenilir puanlamalarda, rubrik kullanmaya ilişkin deneyimin etki etmediği belirlenmiştir. Taştan (2017)'de, Nalbantoğlu Yılmaz ve Gelbal (2011), Alkan (2013) ve Anadol (2017) gibi G kuramını uygulayıp farklı desenleri karşılaştırıp yorumladığı bir çalışma yürütmüştür.

Taştan (2017), G kuramına göre geometri performans görevlerini puanlanmasında öğrencilerin birden fazla puanlayıcı tarafından değerlendirilmesiyle oluşturulan tasarımların G çalışmaları sonuçlarını incelemiştir. Araştırmada 8 görevi, 8 farklı puanlayıcı değerlendirmiştir. ((b:o)xgxp), ((b:o)x(g:p)) ve ((b:(p:o)xg) desenlerinden G çalışmasıyla elde edilen varyanslar ve toplam varyansı açıklama yüzdeleri yorumlanmıştır. ((b:(p:o))xg) desenin varyans bileşenlerini açıklama oranı daha az gibi görünse de G ve Phi katsayıları diğer uygulanan desenlere göre daha büyük elde edilmiştir. 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 sayısının da o kadar artacağı görülmüştür.

Kamış (2018), G kuramında gerçekleştirilen karar çalışmaları için tespit edilen G ve Phi katsayı değerlerinin incelendiği bir araştırma yapmıştır. Araştırmacı, 120 birey, 6 madde ve 12 puanlayıcı kullanarak G kuramını uygulamıştır. Puanlayıcı sayısının artırılması bazı durumlarda G ve Phi katsayılarının artmasına sağlarken bazı durumlarda ise azalmasına sebep olmuştur.

Damlar Demirci (2019), araştırmasında açık uçlu maddelerden oluşmuş matematik sınavının farklı değerlendiricilerin rubrik kullanmadan ve rubrik kullanılarak

puanlanması ile ulaşılan verilere G kuramı uygulayıp karşılaştırmıştır. Çaprazlanmış desen uygulanan çalışmada G ve K çalışmaları yapıp G ve Phi katsayılarını serbest ve anahtarla puanlamaya göre karşılaştırmıştır. Buna göre her iki puanlamanın da güvenilir sonuçlar verdiği görülmüştür. Yapılan K çalışması sonucuna göre puanlayıcı ve madde sayısının artırılması G ve Phi katsayılarının da artmasına sebep olmuştur. Madde sayısının artırılması puanlayıcı sayısının artırılmasına göre daha ekonomik olacağı için madde sayısının artırılması daha güvenilir olacaktır yorumu yapılmıştır.

Başol ve Yüksel (2017), araştırmalarında Fen Bilimleri dersinin bir konusuyla ilgili çoktan seçmeli testi ve yazılı sınavı 8. Sınıf öğrencilerine uygulayıp G ve K çalışması yapmışlardır. Hata varyanslarının yazılı sınavda çoktan seçmeli teste göre yüksek kestirildiğini tespit etmişlerdir. K çalışması ile varyansa etki eden ana sebebin öğrenci değişkeni olduğunu bulmuşlardır.

Yurtiçinde yapılan çalışmalar incelendiğinde, G kuramının, KTK ve Rasch modeli ile karşılaştırılmasının ön planda olduğu aynı zamanda performans değerlendirmede kullanılan ölçeklerin ya da rubriklerin güvenilirliklerinin G kuramına göre değerlendirilmesine yönelik çalışmaların ağırlıkta olduğu görülmektedir. Genellenebilirlik kuramı uygulanarak kurulan desenlerin daha çok çaprazlanmış olarak tasarlanıp ona göre araştırmaların olduğu karışık desenin uygulandığı araştırmaların daha az olduğu görülmüştür.

Yurtdışında Yapılan Araştırmalar

Lane ve Sabers (1989), bireylerin deneysel bir sistemin puanlama güvenilirliğini belirlemek için genellenebilirlik kuramını çalışmalarında kullanmışlardır. Araştırmada bireylerden belirlenmiş bir konu üzerine kompozisyon yazmaları istenmiştir. 15 bireyin kompozisyonlarını 8 puanlayıcı, bir puanlama sistemi oluşturarak değerlendirmiştir. Puanlama sistemi 4 kategoriden oluşmuştur. Çalışmanın amacı, araştırmacıların hazırladıkları puanlama sisteminin güvenilirliğini birçok hata kavramını aynı anda değerlendirebilen genellenebilirlik kuramının çaprazlanmış desenini uygulayarak mutlak ve bağıl hata varyanslarını analiz etmektir. Bu araştırmaya göre puanlayıcı sayısının azalması bağıl ve mutlak hata varyanslarının artmasına ve G ve Phi katsayılarının ise azalmasına sebep olduğu ortaya çıkmıştır.

Shavelson ve diğerleri (1990), Amerikan deniz kuvvetlerindeki erlerin iş performanslarını ölçen görevleri G kuramına göre değerlendirmişlerdir. Sonuçta az

sayıdaki görevden elde edilen verileri işteki performansına genellemenin uygun olmayacağı, güvenilir sonuçlara varabilmek için görev sayısının artırılması dolayısıyla genelleştirebilmenin yapılabileceği yorumunu yapmışlardır.

Goodwin ve Goodwin (1991) çalışmalarında puanlayıcı güvenilirliğini dört farklı yöntemi kullanarak güvenilirlik katsayılarını karşılaştırdıklarında G kuramının okul öncesi eğitim kurumlarındaki ölçme durumlarında güvenilirlik adına birçok açıklama yapılabileceği görülmüştür.

Kunnan (1992), norm referanslı ESL ve EFL testlerinin güvenilirliğini incelemek için G kuramını, faktör ve küme analizini uygulamıştır. Güvenirlik tahminleri, test puanlarının yerleşim puanlarına göre alt testlerinde önemli olduğunu ortaya çıkarmıştır ve yerleşim puanlarının hepsini değerlendirebilmek için küme analizinin uygun olduğu sonucuna varmıştır.

Barnes ve Barnes (1993), çalışmalarında belirli bir kursu alan öğrenciler hakkında kararlar alırken kursu veren öğretmenlerin değerlendirilmesinde genellenebilirlik kuramının pek uygulanmadığını düşünüp çalışmalarında öğretmenleri konu edinmişlerdir. İki adet yuvalanmış desene yer verilmiştir. Birincisi bir öğretmenin birden fazla kurs alanında ders verdiği ikincisi ise bir kursun birden fazla öğretmen tarafında verildiğidir. Bu çalışma için öğretimin organizasyon, kapsam genişliği, grup etkileşimi, coşku, sınıflandırma ve bireysel ilişki adları altında altı boyutu ölçme objesi olarak sınıflandırılmıştır. Birinci örneklem için ölçme objesi öğretmen seçilmiş, ikinci örnekleme ise ölçme objesi kurs seçilmiştir. Sonuçta öğretmenle ilgili kararların belirlenen boyutların herhangi birinden güvenilir kararlar elde edilebileceği fakat kurs kararlarıyla ilgili boyutlar arası büyük farklılıklar ortaya çıkmıştır. Bu çalışmada ölçme objesi birey dışında başka değişkenleri seçen çalışmalara örnek olmuştur.

Sun, Valiga ve Gao (1997), çalışmalarında Cronbach alfa katsayısının akademik danışmanlık öğrencilerinin değerlendirilmelerinin güvenilirliğini değerlendirmek için sıklıkla kullanıldığını fakat bu katsayısının bu tip bireyler hakkında güvenilirlik bilgisi sağlamak için uygun bir yöntem olmadığını belirtmişlerdir. Bu duruma çözüm için 15 kurumun akademik danışmanlık anketinden elde edilmiş veriler kullanılmış ve verilerin analizi için genellenebilirlik kuramı kullanılmıştır. Çalışmanın sonuçlarına göre öğrencilerin danışmanlık notlarının güvenilirliğinin genellenebilirlik kuramına göre daha uygun ve doğru bir şekilde tahmin edilebileceğini göstermiştir.

Lee ve Frisbie (1999), deneysel verileri kullanan test puanlarını iç tutarlılık katsayısı, Cronbach alpha, MTK bilgi parametresi ve G katsayısına göre güvenilirlik

tahmin yöntemlerini karşılaştırmıştır. Güvenirliği tahmin etmede madde puanlarına dayanarak Cronbach alpha kullanılarak aşırı tahmin etmenin büyüklüğü, G teorisi ile yapılan tahmine göre 0.04 yaklaşık çıkmıştır. MTK'na göre yapılan güvenilirlik tahmini ise yerel bağımsızlık göz önünde bulundurularak puanlarının güvenilirliğinin fazla tahmin edilmesiyle sonuçlanmıştır. Aynı çalışmada G katsayısı ve Cronbach alpha katsayıları güven aralıkları açısından karşılaştırıldığında oluşan fark puanların yanlış yorumlanmaya yol açmayacak şekilde olduğu bulunmuştur.

Hagtvet, Frantisek ve Sharma (2001), araştırmalarında test kaygısında özbilgin genellenabilirliği üzerinde çalışmak için bir envanter oluşturmuşlardır. Envanter başarısızlık kaygısını üzerine yuvalanmış özbilgin ve değerlendirilme değişkenleri üzerine kurulmuştur. Envanter Çek Cumhuriyet'inde 766 9. sınıf öğrencisine uygulanmıştır. Tek değişkenli ve çok değişkenli genellenebilirlik kuramı analizleri yapılmıştır. Test kaygısı yaşayan bireylerde dış etkenlerden daha çok etkilendikleri görülmüştür ama genelleştirebilmek adına daha çok kriter üzerinde çalışılması gerektiği öne sürülmüştür.

Gao ve Brennan (2001), çalışmalarında genellenebilirlik kuramını uygulayarak varyans bileşenlerini ve ölçüm hassasiyetini tahmin etme üzerine kavramsal ve istatistiksel çerçeve sağlama amacı gütmüşlerdir. Dinleme ve yazma performansı değerlendirmesi üzerine birkaç yıl içinde toplanan verileri kullanarak tahmini varyans bileşenlerini deneysel olarak araştırmışlardır. G ve Phi katsayılarını yıllar içerisinde değişimlerinin hangi boyutta olduğunu belirlemişlerdir. Ölçümlerin değişmezliğinin yıldan yıla kararlılığını da değerlendirmişlerdir. Sonuçlara göre, tahmin edilen varyans bileşenleri, küçük örneklem gruplarında bir çalışmadan diğerine değişmiştir. G ve Phi katsayıları da yıldan yıla değişiklik göstermiştir. G kuramında uygulanan karar çalışmasının gerçek G çalışması ile elde edilen ölçüm hassasiyetinde olamayacağını belirtmişlerdir.

Hintze ve Matthews (2004), sistematik olarak uygulanan doğrudan gözlemin ortam ve zaman etkisini genelleştirilebilirliğini incelemiştir. 5. Sınıf düzeyindeki 14 öğrenciden oluşan katılımcıların görev içi ve görev dışı davranışları 10 okul günü boyunca günde iki kez anlık bir şekilde doğrudan gözlemlenerek kayda alınmıştır. G kuramı sonuçlarına göre bireylerden kaynaklanan varyansın toplam varyansın %62'sini oluşturduğunu ortaya koymuşlardır. Güvenirlik katsayılarının 0,60'larda çıkmasının sebebi gözlemcilerden kaynaklı farklılıklar olduğunu öne sürmüşlerdir.

Molloy ve Shimura (2005), Japonya'da üniversite çağındaki öğrenciler İngilizce öğreniminde on iki yazılı şikayette bulunmuşlardır. Üç veri seti olarak incelemiştir.

Verileri üzerinde G ve K çalışması yapmıştır. Kestirilen üç veri setindeki varyansın bireysel farklılıklarla ilgili olduğunu belirtmişlerdir.

Wang (2005), araştırmasında KTK ve G kuramını karşılaştırmıştır. Bu amaçla İngilizce dersi yazılı anlatım becerisini ölçen bir testin verilerini kullanarak maddelerin becerilere yuvalandığı (p x i : t) çaprazlanmış ve yuvalanmış desenin birlikte uygulandığı deseni kullanmıştır. G çalışması sonuçlarına göre bireylere bağlı varyasyonun büyük olduğu görülmüştür. Bunun nedeni birey etkisinin varyasyonunun madde etkisi varyasyonun iki katı olduğudur. Yazma yeteneğindeki değişim bireylerden kaynaklanmakta olduğu yorumu yapılmıştır.

Bruckner, Yoder ve Mcwilliam (2006), dilbilgisi ve fonolojik bozuklukları olan okul öncesi çocuklardan elde edilen konuşma dili örnekleri kullanılarak iki yüzeyle çaprazlanmış bir desen uygulanarak G çalışması yapmışlardır. Ortalama ifade uzunluğu (MLU), farklı kelime kökleri sayısı ve tamamen anlaşılabilir ifade girişimleri adı altında üç değişken ölçülmüştür. Bulgulara göre toplam varyans için en yüksek varyans değeri birey ve puanlayıcı ortak etkisinden kaynaklandığı görülmüştür. Bir uygulayıcı ve bir uygulama seansı ile 0.70'e eşit ya da büyük G katsayısı değerlerine ulaşılmıştır. K çalışması sonuçlarına göre yüksek güvenilir sonuçlar elde edebilmek için uygulayıcının en az beş uygulama seansı yapması gerektiği belirtilmiştir.

Lee ve Kantor (2007), çalışmalarında TOEFL sınavının yazma becerileri bölümü için çeşitli derecelendirme tasarımlarını ve görev ve puanlayıcı sayılarının etkisini incelemişlerdir. Yazma becerilerinin ölçüm güvenilirliğini tek değişkenli ve çok değişkenli G kuramına göre analiz etmişlerdir. Analiz sonuçlarına göre görev sayısını artırmak yerine puanlayıcı sayısını artırmanın daha güvenilir sonuçlar vereceği sonucuna ulaşmışlardır. Çok değişkenli tasarımlarla tek değişkenli tasarımlara göre nispeten daha yüksek puan güvenilirliği elde etmişlerdir.

Lakes ve Hoyt (2009), çocuk ve ergenlik çağı önlemlerinin güvenilirliğini değerlendirmek için genellenebilirlik kuramının kullanılmasının araştırmacılara daha hassas ölçümler yapabilmeyi ve ölçme sonuçlarına olan güveni artırdığını sağladığı fikri ile yola çıkan araştırmacılar belirledikleri çeşitli değerlendirmeleri kullanarak 181 çocuk üzerinde öğretmen ya da veli değişkenini de çalışmalarına katarak G ve K çalışmasının önemini ortaya koymaya çalışmışlardır.

Wasserman, Levy ve Loken (2009), psikoterapi süreci çalışmalarında ortak olan kodlama prosedürleri (ICC), psikoterapi süreci ölçüm geliştirme sürecinde güvenilirliği doğru değerlendirilmesi psikoterapiye etki eden mutatif değişkenler bakımından çok

yönlü olarak analiz edilmesi amacıyla G kuramını uygulamışlardır. Borderline kişilik bozukluğu için psikoterapi süreci değerlendirme ölçeğindeki tekniklerin geliştirilebileceği üzerine çalışmışlardır.

Alhaktani (2012), Kuran okunuşuna yönelik hazırlanmış dereceli puanlama anahtarını G kuramı, Rasch ölçüm modeline göre incelemiştir. 93 öğrenci 3 puanlayıcı tarafından bütünsel ve analitik olarak derecelendirilmiştir. G kuramı sonuçlarına göre madde ve puanlayıcıların artırılması güvenilirliği artırmıştır ama puanlayıcı sayısının artırılması güvenilirliği daha çok artırmıştır. Zaman ve emek düşünüldüğünde bu farklılığın önemsenmeyeceği düşünülmüştür.

Gugiu ve Gugiu (2012), araştırmalarında lisans öğrencilerinin yazılı ödevlerinin değerlendirilmesinde elde edilen puanların güvenilirliğini genellenebilirlik kuramına göre test etmişlerdir. Bu amaçla iki puanlayıcı ve dört farklı derecelendirilmiş rubrik kullanılan 9 grup oluşturulmuştur. Toplam 72 veriden oluşan araştırmada çaprazlanmış desen uygulanarak G ve D çalışmaları yapılmıştır. Sonuçta iki farklı puanlayıcının performans değerlendirmeleri arasında dikkat çekici farklılıklar olduğu ortaya çıkmıştır.

Hintze, Owen, Shapiro ve Daly III (2012), araştırmalarını sözlü okuma akıcılığını müfredata dayalı ölçme yöntemi kullanarak ölçmüşlerdir ve bu ölçümlerini KTK'ye alternatif olarak genellenebilirlik kuramına göre değerlendirip hata kaynaklarının önlenmesinde ya da belirlenmesinde uygun bir yöntem olduğu bulunmuştur.

Bottema-Beutel, Lloyd, Carter ve Asmus (2014), davranışın birden fazla birbiriyle ilişkili faktörden etkilenebileceği düşüncesiyle okul ve sınıf ortamında güvenilir tahminlere genellenebilirlik kuramı uygulayarak ulaşmaya çalışmışlardır. Kapsayıcı ortaöğretim sınıflarında ağır engelli öğrencilerin sosyal ve akademik katılımına odaklanan rastgele bir kontrol çalışmasının gözlemsel verileriyle G ve K çalışmaları yapmışlardır. Elde ettikleri sonuçlarda gözlemsel kararların mevcut veri setlerinden gelen güvenilirlik tahminlerine göre yapılmasının önemli olduğu ortaya çıkmıştır. G ve K çalışmalarına göre güvenilirlik tahminleri istenenden düşük kestirilmiştir.

Lin ve Zhang (2014), İngilizce dil yeterlik standartları ile akademik içerik standartları arasındaki ilişki üzerine yapılan çalışmalarla ilgili akademik disiplinlerle ilişkili dil performans göstergeleri hakkındaki puanlayıcıların değerlendirmelerinin güvenilirliğini genellenebilirlik kuramına göre incelemiştir. İçerik uzmanı ve ESL uzmanından oluşan 500'ün üzerinde puanlayıcıdan elde edilen 20 mektup ile çalışılmıştır. Değişkenlik kaynakları olarak puanlayıcı güvenilirliğini ve standart ölçüm hatalarını belirlemiştir. Sonuçlara göre, belirli sınıf ve konu alanlarına bağlı olarak, kabul

edilebilir güvenilirlik seviyesine ulaşmak ve puanlamalardan ölçüm hatalarını kontrol etmek için 3-6 puanlayıcının puanlamalarının olması gerektiğine ulaşılmıştır.

Han (2018), akademik olarak derecelendirme ve sertifikalandırma gibi yüksek riskli karar alma süreçlerini içeren değerlendirmeler ile ilgili sınırlı deneysel çalışmaların olduğu fikrinden yola çıkarak çevirmenlik kurslarındaki puanlayıcıların puanlamalarına birden fazla desen uygulayarak genellenebilirlik kuramını uygulamıştır. Bu çalışma ile İngilizce ve Çince dillerinin çevirmenliğinde puanlayıcı güvenilirliğini elde edebilmek için optimal ölçüm tasarımını keşfetmeyi amaçlamıştır. Bulgulara göre puanlayıcılar, aktarımın akıcılığı ve hedef dil kalitesi ölçeklerinden ziyade bilgi tamlığı ölçeğini kullanarak daha tutarlı bulunmuşlardır. Puanlayıcıların, İngilizce'den Çince'ye yorumlamaların, Çince'den İngilizce'ye yorumlamalarında daha tutarlı oldukları sonucuna ulaşmıştır.

Ohta, Plakans ve Gebril (2018), çalışmalarında okuma yazma maddelerinden (RTW) elde edilen puanların geliştirilebilirliğini araştırmışlardır. Bu amaçla hem bütünsel hem de çok özellikli derecelendirme ölçeklerinden alınan puanların güvenilirliğini G kuramına göre karşılaştırmışlardır. 60 EFL üniversitesi öğrencisinin yazdığı 60 makaleler bütünsel ve çok özellikli değerlendirme ölçeklerine göre 5 puanlayıcı tarafından değerlendirilmiştir. Sonuçlara göre çok özellikli derecelendirme ölçeğinin, bütünsel değerlendirme ölçeğini göre daha güvenilir olduğu sonucuna varmışlardır.

BÖLÜM 2

YÖNTEM

Araştırmanın türü ve çalışma grubuna ait bilgilendirmelerin yer aldığı bölümdür. Aynı zamanda araştırma verileri ve verilerin çözümlenmesinde uygulanan analiz yöntemlerinden bahsedilen açıklamalar bulunmaktadır.

Araştırmanın Modeli

Birey ya da grubun özelliklerini olduğu gibi tanımlayıp ortaya koyan, bu durumu nicel ya da nitel açılardan tanımlayan aynı zamanda ileriki araştırmalar için fikir üretmeye yönelik öngörü sağlayan araştırma türü betimsel araştırmadır. Betimsel araştırmalar, tarama modelinde ele alınabilmektedir (Karasar, 1998; Erkuş, 2009). Bu araştırma, personel alımı sınavının farklı puanlayıcılar tarafından değerlendirilen sonuçlarının genellenebilirlik kuramına göre incelenmesini ve karar çalışması yapılmasını konu edindiği için tarama modelinde betimsel bir araştırmadır.

Çalışma Grubu

Araştırmada yer alan çalışma grubu, ASYM' de açık uçlu maddelerden oluşan bir sınava katılan 211 bireyin personel alımı sınavındaki Hukuk alt testinde yer alan maddelerin iki puanlayıcı ile değerlendirilmesi sonucunda elde edilen verilerden oluşmuştur. Puanlayıcılar değerlendirilmelerinde cevaplama anahtarı kullanmışlardır.

Verilerin Toplanması

Araştırmada, Ankara Üniversitesi Sınav Yönetim Merkezinden (ASYM) resmi izinle alınan personel alımı sınavının Hukuk alt testine ait açık uçlu maddelerden elde edilen veriler kullanılmıştır.

Testte bireylere dört soru sorulmuştur ve bu sorulardan istedikleri iki tanesini cevaplamaları istenmiştir. Bireylerin cevapları da iki puanlayıcı tarafından

değerlendirilmiştir. Her birey bütün soruları cevaplamamıştır ve iki puanlayıcı tarafından değerlendirilmiştir. Elde edilen verilere G kuramına dayalı olarak G ve K çalışması yapılmıştır.

Verilerin Çözümlemesi

Verilerin çözümü için genellenabilirlik kuramına göre belirlenen ($b \times m \times p$) ve ($b \times (m:p)$) desenlerin ayrı ayrı analizi yapıp elde edilen sonuçlar karşılaştırılmıştır.

Açık uçlu maddelerin hangilerinin, hangi birey tarafından yanıtladığının ve puanlayıcıların hangi maddeleri puanladığı bilgisinin listelemesini yapabilmek için Excel paket programı kullanılarak veriler analize hazırlanmıştır. Bireylerin en çok seçtikleri veriler sınırlandırılarak 93 birey ile oluşturulmuş veri üzerinde ($b \times m \times p$) çaprazlanmış desen uygulanmıştır. Yine veriler sınırlandırılarak 40 veri üzerinde bireylerin maddelerle ve puanlayıcılarla çapraz, maddelerin puanlayıcılara yuvalandığı ($b \times (m:p)$) yuvalanmış desen uygulanmıştır.

Araştırmanın alt amaçlarındaki soruların yanıtlanabilmesi için iki puanlayıcı ölçme durumları oluşturulup çaprazlanmış desen ve yuvalanmış desenin uygulandığı analizde göreceli ve mutlak hata varyansları ve G ve Phi katsayı değerleri elde edilmiştir. İki puanlayıcı ölçme durumları için K çalışması yapılmıştır. İki farklı desene göre yapılan G ve K çalışmalarının sonuçları karşılaştırılıp yorumlanmıştır.

G kuramında, oluşturulan desenlere yönelik varyans bileşenlerinin belirlenmesi, belirlenen bileşenlerin toplam varyansa etki ettiği yüzde değerlerinin kestirilmesi ve karar çalışmalarının yürütülmesinde EduG programı kullanılmıştır (Cardinet, Johnson ve Pini 2010).

BÖLÜM 3

BULGULAR VE YORUMLAR

Araştırmanın alt amaçlarının ışığında verilerin çözümlenmesi ile ulaşılan değerlere ve elde edilen bu değerlere göre yapılan yorumlar bulunmaktadır. Uygulanan iki desen ile kestirilen varyans bileşenleri ve bu bileşenlerin değerlerinin bütün analiz içindeki varyans yüzdelerini gösteren grafiklerde bu bölümde yer almaktadır.

Birey (b), Madde (m) Ve Puanlayıcı (p) Değişkenlerinin Çapraz Tasarlandığı (bxm xp) Desenin G Kuramı Sonuçları

Araştırmanın birinci alt amacı olan (bxm xp) çaprazlanmış desenine ait G ve K çalışmaları ayrı başlıklarda incelenmiştir.

(bxm xp) Desenin G Çalışması Sonucu Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Bu başlıkta veriler sınırlandırılarak elde edilmiş 93 bireyin cevaplarının 2 puanlayıcı tarafından değerlendirildiği (bxm xp) çaprazlanmış desende kullanılan yüzeylerin sayısal değerleri, varyans bileşenleri ve bu bileşenlerin yorumlanması, G ve Phi katsayı değerleri ve değerlerin yorumlanması ile ilgili bilgiler bulunmaktadır. Aşağıdaki tablo 3'te veriyi oluşturan yüzeyler ve bu yüzeylerin seviyeleri ile ilgili bilgi verilmiştir.

Tablo 3

Desende Kullanılan Yüzeylerin Sayısal Değerleri

Facet (Yüzey)	Seviye
Birey	93
Madde	2
Puanlayıcı	2

İşe giriş sınavında uygulanan, toplam 4 maddeden 2 tanesinin seçileceği bir ölçme kuralına göre cevaplandırılacak sınavda 211 birey içinden 93'ünün 1 ve 4 numaralı maddeyi seçtiği görülmüştür. Buna göre bu maddeler, bireyler tarafından en çok seçilen maddeler olmuştur. Seçilen maddeler veri içerisinde sınırlandırılarak uygulayacağımız desene hazır hale getirilmiştir. Bireylerin bütün maddeleri cevapladığı ve maddelerin bütün puanlayıcılar tarafından değerlendirildiği çaprazlanmış desenin G çalışması sonuçları Tablo 4'tedir.

Tablo 4

(bxm xp) Deseni Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

Varyans Kaynağı	Kareler Toplamı (SS)	Sd	Kareler Ortalaması (MS)	Varyans	%
b	33478,28	92	363,89	30,30	16,6
m	27,42	1	27,42	0,02	0,00
p	5351,94	1	5351,94	27,41	15,00
bm	811,33	92	8,82	1,20	0,70
bp	22107,31	92	240,30	116,94	64,10
mp	20,35	1	20,35	0,15	0,10
bmp, e	590,90	92	6,42	6,42	3,50
Toplam	62388,03	371			100

b: Birey, m: Madde, p: Puanlayıcı

Tablo 4 incelendiğinde birey (b) ana etkisi için kestirilen varyans bileşeni toplam varyansın %16,6'sını oluşturduğu görülmektedir. Bu sonuçla bireylerin değişkenlik yüzdelerinin olduğu ama büyük bir değer olmadığı görülmüştür. Bu tip bir sınava girebilmek için belirli bir başarı düzeyinin üzerinde olmak gerekmektedir. Bu nedenle bireyler arasında farklılaşmanın fazla olmaması olağan kabul edilebilmektedir. Puanlayıcıların toplam varyans içindeki etkisinin fazla kestirilmesi de birey ana etkisinin olması gerekenden düşük kestirilmesine yol açmış olabileceği düşünülmektedir. Aynı zamanda madde sayısının az olması da bireyler arası farklılık bu oranda kestirilmiş olabilir. Madde (m) ana etkisi için kestirilen varyans bileşeninin toplam varyansa etkisi

olmadığı görülmüştür. Madde ana etkisinin sıfır çıkması maddelerin zorluk derecesinin değişkenlik göstermediğinin bir belirtisi yorumu yapılabilir. Aynı zamanda madde sayısının sadece iki olması da madde etkisinin toplam varyansa etki etmemesine yol açmış olabilir. Puanlayıcı ana etkisi bazı puanlayıcıların öğrencileri değerlendirmede diğer puanlayıcılara göre ne kadar farklılaştığı ile ilgili bilgi verir. Sonuçlara göre puanlayıcı (p) ana etkisi için elde edilen varyans bileşeninin toplam varyansa etkisinin %15 çıktığı görülmüştür. Bu değer incelendiğinde puanlayıcılar arası farklılık olduğunu, puanlayıcıların bireylerin sonuçlarını değerlendirirken farklı yorumlarda bulduklarını göstermektedir.

Birey x Madde (bm) ortak etkileşimine ait varyans bileşeni toplam varyansın %0,7'sini açıklayarak kestirilen en küçük üçüncü varyans değerine sahiptir. Bu durum bireylerin maddeden maddeye farklılık oluşturacak bir değişiklik yaratmadığını göstermektedir. Madde ana etkisinin varyans bileşeni "0" a yakın çıktığı için birey x madde etkisi sonucunun da değişkenlik göstermeyecek şekilde çıkması analizi desteklemektedir. Bu sonuç madde sayısının az olması ve puanlayıcı ana etkisinin fazla olmasıyla açıklanabilmektedir. Birey x Puanlayıcı (bp) ortak etkileşimi için kestirilen varyans bileşeni toplam varyansın %64,1'ini açıklayarak kestirilen en büyük varyans değerine sahip olduğu görülmüştür. Puanlayıcılar bireyleri puanlarken bireyden bireye farklılıklar oluşturmuşlardır. Puanlayıcıların değerlendirmesi bireyden bireye değişkenlik göstermiş dolayısıyla bireyleri puanlamada tutarsız olduklarını ifade etmektedir. Bu durum puanlayıcı ana etkisi için çıkan sonuç ile de desteklenmektedir. Madde x Puanlayıcı (mp) ortak etkileşimi için kestirilen varyans bileşeni toplam varyansın %0,1'ini açıklayarak kestirilen en küçük ikinci varyans değerine sahiptir. Bu durum puanlayıcıların maddeden maddeye farklılık oluşturmayacak şekilde puanlama yaptıklarını ifade etmektedir. Birey x Maddex Puanlayıcı (bmp) ortak etkileşimi için kestirilen varyans artık varyanstır. Artık varyans bileşenine ait varyans toplam varyansın %3,5'ini oluşturmaktadır. Artık bileşeninden kaynaklanan bu varyans, birey x madde x puanlayıcı ortak etkileşiminde tesadüfi hata kaynaklarının varlığına işaret eden bir değerdir. Artık varyans bileşenin olabildiğince düşük olması istenir. Bu analizde de artık bileşenden kaynaklı bir değişkenlik olmuştur ve %3,5 değerinde bir varyasyonun kaynağı bilinmediği için tesadüfi hatalar vardır. Artık varyansa ait değer puanlayıcı ana etkisinden kaynaklanan bir varyans varlığı tespit edildiği için daha düşük kestirilmiştir.

Tablo 5

(bxm_{xp}) *Desenine Ait G ve Phi Katsayı Değerleri*

Katsayılar	Değer
G katsayısı	0,33
Phi katsayısı	0,29

Tümüyle çaprazlanmış desen için kestirilen Tablo 5'teki G ve Phi katsayıları incelendiğinde bağıl hata varyansına yönelik hesaplanan G katsayısının 0,33, mutlak hata varyansına yönelik hesaplanan Phi katsayısının ise 0,29 çıktığı görülmektedir. Bu sonuçlara göre elde edilen güvenilirlik katsayıları oldukça düşük çıkmıştır. Uygulanan testin bireyler ya da puanlayıcılar arası farklılığı iyi yorumlayamadığı düşünülmektedir. Elde edilen düşük G ve Phi katsayı değerlerinin puanlayıcıların bireyleri değerlendirmelerinde tutarlı olmamalarından kaynaklandığı düşünülmektedir.

(bxm_{xp}) Deseninde Madde ve Puanlayıcı Sayılarının Artırılması ve Azaltılmasıyla Yapılan K Çalışması

93 bireyin dört madde içerisinde iki madde seçmesi ve bu maddelerin iki puanlayıcı tarafından değerlendirilmesinden oluşan veriye K çalışması uygulanmıştır. (bxm_{xp}) deseninde bireyler ölçme objesi olarak belirlenip madde ve puanlayıcı sayılarının artırılması ve azaltılmasıyla oluşturulan durumlar için kestirilen G ve Phi katsayısı, göreceli ve mutlak hata varyanslarına ait sonuçlar Tablo 6'da bulunmaktadır.

Tablo 6

(bxm_{xp}) *Deseni Madde ve Puanlayıcı Sayılarının Değiştirilmesiyle Kestirilen G ve Phi Katsayı Değerleri, Bağıl ve Mutlak Hata varyansları*

N_M	N_P	G	Phi	$\sigma^2(\delta)$	$\sigma^2(A)$
2	2	0,33	0,29	60,67	74,42
2	3	0,43	0,38	40,64	49,82
3	2	0,34	0,29	59,94	73,68
2	1	0,20	0,17	120,75	148,24

(Devam ediyor)

Tablo 6 (devam)

(bxm_{xp}) *Deseni Madde ve Puanlayıcı Sayılarının Değiştirilmesiyle Kestirilen G ve Phi Katsayı Değerleri, Bağlı ve Mutlak Hata varyansları*

N_M	N_P	G	Phi	$\sigma^2(\delta)$	$\sigma^2(A)$
1	2	0,32	0,28	62,88	76,68
4	2	0,34	0,29	59,57	73,30
6	2	0,34	0,29	59,20	72,92
2	7	0,63	0,58	17,76	21,70
2	8	0,66	0,61	15,62	19,07
2	10	0,71	0,66	12,62	15,38
2	12	0,74	0,70	10,61	12,92
3	3	0,43	0,38	40,09	49,25
6	6	0,60	0,55	19,87	24,44

N_M : Madde Sayısı, N_P : Puanlayıcı Sayısı

Araştırmada kullanılan 2 puanlayıcının 2 maddeyi puanlamasına göre yapılan K çalışmasında G katsayısı 0,33 kestirilmiştir. Tablo 6'daki değerlere göre madde sayısı ($N_M = 2$) sabit iken puanlayıcı sayısı ($N_P = 1$) azaltıldığında G katsayısının 0,20 ile gerçek durumdaki değerinden düşük çıktığı görülmüştür. Puanlayıcı sayısı sabit tutulup ($N_P = 2$) madde sayısı artırıldığında ($3 \leq N_M \leq 6$) ise G katsayısının kayda değer bir değişkenlik göstermediği görülmüştür. Madde sayısı ($N_M = 2$), puanlayıcı sayısı ($N_P = 2$) olduğunda 0,33 olan G katsayısı, madde sayısı ($N_M = 6$), puanlayıcı sayısı ($N_P = 2$) olduğunda elde edilen G katsayısı 0,34 çıkmıştır. Madde sayısı sabit tutulup ($N_M = 2$) puanlayıcı sayısı artırıldığında ($3 \leq N_P \leq 12$) ise G katsayısının güvenilir sayılabilecek düzeylere ulaştığı görülmektedir. Madde sayısı ($N_M = 2$), puanlayıcı sayısı ($N_P = 12$) olarak belirlendiğinde G katsayısı 0,74 olmuştur. Elde edilen G katsayısı ile veriyi oluşturan ($N_M = 2$, $N_P = 2$) ilk duruma göre G katsayısında artış 0,41 olmuştur. Puanlayıcı ve madde sayısının her ikisinin de artırıldığı durumda ise G katsayısı ($N_M = 3$, $N_P = 3$) olduğunda verinin ilk halinden elde edilen ($N_M = 2$, $N_P = 2$) katsayıya göre 0,33'ten 0,43'e yükseldiği, ($N_M = 6$, $N_P = 6$) olduğunda ise 0,33'ten 0,60'a yükseldiği görülmüştür.

Araştırmada kullanılan 2 puanlayıcının 2 maddeyi puanlamasına göre yapılan K çalışmasında Phi katsayısı 0,29 kestirilmiştir. Tablo 6'daki değerlere göre madde sayısı ($N_M = 2$) sabit iken puanlayıcı sayısı ($N_P = 1$) azaltıldığında Phi katsayısının 0,17 ile gerçek

durumdaki değerinden düşük çıktığı görülmüştür. Puanlayıcı sayısı sabit tutulup ($N_P = 2$) madde sayısı artırıldığında ($3 \leq N_M \leq 6$) ise Phi katsayısının anlamlı bir değişkenlik göstermediği görülmüştür. Örneğin, madde sayısı ($N_M = 2$), puanlayıcı sayısı ($N_P = 2$) olduğunda 0,29 olan Phi katsayısı, madde sayısı ($N_M = 6$), puanlayıcı sayısı ($N_P = 2$) olduğunda elde edilen Phi katsayısı 0,29 çıkmıştır. Madde sayısı sabit tutulup ($N_M = 2$) puanlayıcı sayısı artırıldığında ($3 \leq N_P \leq 12$) ise Phi katsayısının güvenilir denilebilecek düzeylere ulaştığı görülmektedir. Madde sayısı ($N_M = 2$), puanlayıcı sayısı ($N_P = 12$), olarak belirlendiğinde Phi katsayısı 0,70 olmuştur. Elde edilen Phi katsayısı ile veriyi oluşturan ($N_M = 2, N_P = 2$) ilk duruma göre Phi katsayısında artış 0,41 olmuştur. Puanlayıcı ve madde sayısının her ikisinin de artırıldığı durumda ise Phi katsayısı ($N_M = 3, N_P = 3$) olduğunda verinin ilk halinden elde edilen ($N_M = 2, N_P = 2$) katsayıya göre 0,29'dan 0,38'e yükseldiği, ($N_M = 6, N_P = 6$) olduğunda ise 0,29'dan 0,55'e yükseldiği görülmüştür.

Tablo 6'da madde sayısı sabit tutulup ($N_M = 2$), puanlayıcı sayısının artırılması ($3 \leq N_P \leq 12$) durumunda bağıl ($\sigma^2(\delta)$) ve mutlak ($\sigma^2(\Delta)$) hata varyansları incelendiğinde, puanlayıcı sayısındaki artışa bağlı olarak hata varyanslarının azaldığı görülmektedir. Örneğin, madde sayısı ($N_M = 2$), puanlayıcı sayısı ($N_P = 2$) olduğu durumda bağıl ve mutlak hata varyansları ($\sigma^2(\delta) = 60,67$), ($\sigma^2(\Delta) = 74,42$) iken madde sayısı ($N_M = 2$), puanlayıcı sayısı ($N_P = 12$) olduğu durumda bağıl ve mutlak hata varyansları ($\sigma^2(\delta) = 10,61$), ($\sigma^2(\Delta) = 12,92$) ile bağıl hata varyansında 50,06, mutlak hata varyansında ise 72,5 değerinde bir fark olduğu görülmektedir. Madde sayısı sabit iken puanlayıcı sayısının artırılmasıyla elde edilen hata varyanslarındaki bu azalmanın G ve Phi katsayılarındaki artış ile de desteklendiği görülmektedir. Madde sayısı sabit tutulup ($N_M = 2$), puanlayıcı sayısının azaltılması ($N_P = 1$) durumunda bağıl ve mutlak hata varyansları ($\sigma^2(\delta) = 120,75$), ($\sigma^2(\Delta) = 148,24$) madde ve puanlayıcı sayısının ($N_M = 2, N_P = 2$) olduğu durumda elde edilen bağıl ve hata varyanslarına göre ($\sigma^2(\delta) = 60,67$), ($\sigma^2(\Delta) = 74,42$) bağıl hata varyansı 60,08, mutlak hata varyansı 73,82 değerinde bir artış olduğu görülmüştür. Puanlayıcı sayısı sabit tutulup ($N_P = 2$), madde sayısı artırıldığında ($3 \leq N_M \leq 6$) bağıl ve mutlak hata varyansları incelendiğinde ise madde sayısındaki artışa bağlı olarak hata varyansları önemli bir değişiklik göstermemiştir. Örneğin puanlayıcı sayısı ($N_P = 2$), madde sayısı ($N_M = 2$) olduğu durumda bağıl ve mutlak hata varyansları ($\sigma^2(\delta) = 60,67$), ($\sigma^2(\Delta) = 74,42$) iken puanlayıcı sayısı ($N_P = 2$), madde sayısı ($N_M = 6$) olduğu durumda bağıl ve mutlak hata varyansları ($\sigma^2(\delta) = 59,20$), ($\sigma^2(\Delta) = 72,92$) ile bağıl hata varyansında 1,47, mutlak hata varyansında 1,55 gibi küçük bir farkın olduğu görülmektedir.

Puanlayıcı ve madde sayısının her ikisinin de artırıldığı durumda ise bağıl ve mutlak hata varyansları da düşüş göstermektedir. Örneğin, madde ve puanlayıcı sayısı ($N_M = 6, N_P = 6$) olduğunda bağıl ve mutlak hata varyansı ($\sigma^2(\delta) = 19,87$), ($\sigma^2(\Delta) = 24,44$), madde ve puanlayıcı sayısının ($N_M = 2, N_P = 2$) olduğu durumda elde edilen bağıl ve hata varyanslarına göre ($\sigma^2(\delta) = 60,67$), ($\sigma^2(\Delta) = 74,42$) bağıl hata varyansında 40,8, mutlak hata varyansında ise 49,98 bir düşüş olmuştur.

Yukarıda yer alan analizlere göre puanlayıcı sayısı artırıldığında güvenilirliğin de arttığı görülmüştür fakat puanlayıcı sayısı sabit tutulup madde sayısı artırıldığında güvenilirliğin artmadığı gözlemlenmiştir. Bunun yanı sıra madde ve puanlayıcı sayılarının aynı anda artırılması da güvenilirliği artırabilmektedir. Testte kullanılan maddelerin bireyler arası farklılığı iyi yordayan maddeler olması sağlanmalıdır. Puanlayıcı ve madde sayısının aynı anda artırılması güvenirlığe olumlu katkı sağladığı için güvenilirliği artırmak amacıyla her iki değişkenin de sayısının artırılmasının uygun olacağı düşünülmektedir. Fakat bu madde ve puanlayıcı sayısı artırılırken bazı önemli faktörlere dikkat edilmelidir. Örneğin testte kullanılacak olan maddelerin bireyler arası farklılığı iyi yordayan maddeler olması sağlanmalıdır. Ayrıca bireyleri değerlendirecek puanlayıcıların mutlaka dereceli puanlama anahtarı kullanmaları gerekmektedir. Bütün faktörler göz önünde bulundurularak iş gücü ve zaman açısından olumsuz olmayacak şekilde puanlayıcı ve madde sayısının aynı anda artırılacağı açık uçlu bir sınav düzenlenmesinin daha avantajlı olacağı öngörülmektedir.

Birey (b) değişkenin çapraz, Madde (m) Ve Puanlayıcı (p) Değişkenlerinin Yuvalanmış Tasarlandığı (bx(m:p)) Deseninin G Kuramı Sonuçları

Araştırmanın ikinci alt amacı olan (bx(m:p)) çaprazlanmış desenine ait G ve K çalışmaları ayrı başlıklarda incelenmiştir.

(bx(m:p)) Deseninin G Çalışması Sonucu Kestirilen Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

İşe giriş sınavında tercihli maddelerden oluşan testte 3. ve 4. maddeleri seçen 40 bireyin cevapları farklı puanlayıcılar tarafından değerlendirilmiştir ancak puanlayıcılar farklı maddeleri puanlamıştır. Dolayısıyla bireyler bütün puanlayıcılar tarafından değerlendirilmiş ama bütün maddeler bütün puanlayıcılar tarafından değerlendirilmediği

için bireyler puanlayıcılar ve maddeler ile çaprazlanarak, puanlayıcılar ile maddeler yuvalanarak (bx(m:p)) deseni uygulanmıştır. (bx(m:p)) deseninde madde ve puanlayıcı olmak üzere iki yüzey yer almaktadır. Birey ise ölçme objesidir. (bx(m:p)) desenli ölçmede bulunan değişkenlik kaynakları birey (b), puanlayıcı (p), maddelerin puanlayıcılara yuvalandığı (m:p), bireylerin puanlayıcılar ile çaprazlandığı (bp) ve artık varyans (bx(m : p), e) olmak üzere 5 adettir.

Yuvalanmış desende yüzeylerin sayısal değerleri, varyans bileşenleri ve bu bileşenlerin yorumlanması, G ve Phi katsayı değerleri ve bu değerlerin yorumlanması ilgili bilgiler bulunmaktadır. Aşağıdaki tabloda veriyi oluşturan yüzeyler ve bu yüzeylerin seviyeleri ile ilgili bilgi verilmiştir.

Tablo 7

Desende Kullanılan Yüzeylerin Sayısal Değerleri

Facet (Yüzey)	Seviye
Birey	40
Puanlayıcı	2
Madde içinde puanlayıcı	2

İşe giriş sınavında uygulanan, toplam 4 maddeden 2 tanesinin seçileceği bir ölçme kuralına göre cevaplandırılacak sınavda 40 birey 3. ve 4. maddeleri cevaplandırmıştır. Bu maddeler farklı puanlayıcılar tarafından değerlendirilmiştir. Bireylerin aynı maddeleri cevaplandırarak farklı puanlayıcılar tarafından farklı maddelerin seçilerek değerlendirilmesi ile oluşturulmuş yuvalanmış desenin G çalışması sonuçları Tablo 8’de verilmiştir.

Tablo 8

(bx(m:p)) Deseni Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri

<i>Varyans Kaynağı</i>	<i>Kareler Toplamı (S.S.)</i>	<i>Sd</i>	<i>Kareler Ortalaması (M.S.)</i>	<i>Varyans</i>	<i>Varyans Yüzdesi</i>
b	11454,99	39	293,72	30,5	16,24
p	1494,51	1	1494,51	0,0	29,36

(Devam ediyor)

Tablo 8 (devam)

(bx(m:p)) *Deseni Varyans Bileşenleri ve Toplam Varyansı Açıklama Yüzdeleri*

<i>Varyans Kaynağı</i>	<i>Kareler Toplamı (S.S.)</i>	<i>Sd</i>	<i>Kareler Ortalaması (M.S.)</i>	<i>Varyans</i>	<i>Varyans Yüzdesi</i>
m:p	5676,61	2	2838,31	30,2	50,18
bp	625,24	39	16,03	0,0	7,31
b(m:p),e	6994,89	78	89,68	39,4	14,18
Toplam	26246,24	159			100

b: Birey, m: Madde, p: Puanlayıcı

Tablo 8 incelendiğinde kestirilen birey (b) ana etkisi varyans yüzdesinin %16,24 olduğu görülmektedir. Ölçme objesi olan bireylere ait varyans bileşeni yüzde değerinin bireyler arası farklılıkları ortaya çıkarabilmesi adına yüksek olması beklenir. Bu sonuçla bireylerin değişkenlik yüzdelerinin üçüncü büyük varyans yüzdesi değerine sahip olduğu görülmüştür. Böyle bir sınava girebilmek için belirli bir başarı düzeyinin üzerinde olmak gerektiğinden bireyler arası farklılıkların çok yüksek olmaması olağan kabul edilebilmektedir.

Puanlayıcı (p) ana etkisi için kestirilen varyans bileşeninin varyans yüzdesi %29,36 ile ikinci en büyük varyans açıklama yüzdesine sahiptir. Bu değer incelendiğinde puanlayıcılar arası farklılık olduğunu, puanlayıcıların bireylerin sonuçlarını değerlendirmelerinde farklılıklar bulunduğu yorumu yapılabilir.

Madde puanlayıcı (m:p) etkisine ait varyans toplam varyansın %50,18'ini açıklayarak en yüksek varyans değerine sahiptir. Her bir puanlayıcının puanladığı maddelere ilişkin bireylerin cevaplarında farklılıklar oluşmuştur. Maddelerin günlük düzeyleri arasındaki farklılıklar, tüm maddeleri tüm puanlayıcılar puanlamadığı için maddeler hakkında puanlayıcılardan bağımsız yorum yapılamaz. Puanlayıcıların tutarlılıkları maddeden maddeye değişkenlik göstermiştir.

Birey puanlayıcı (bp) ortak etkileşimine ait varyans toplam varyansın %7,31 ini açıklayarak en düşük varyans değerine sahiptir. Puanlayıcıların bireyleri puanlamada az da olsa tutarsızlıkları olduğunu belirtir. Puanlayıcılar değerlendirmelerinde bireyden bireye farklılıklar göstermiştir. Puanlayıcı ana etkisi için kestirilen varyans değeri ile de bu sonuç desteklenmektedir.

Artık varyans bileşeni toplam varyansın %14,18'ini açıklamaktadır. Artık

bileşenden kaynaklanan bu varyans değeri, maddeler ve puanlayıcılar arasındaki etkileşim etkisi ile bu çalışmada ölçülememiş olabilen tesadüfi değişkenlik kaynaklarının varlığının bir göstergesidir. Artık varyans bileşeninin değerinin oldukça düşük olması arzu edilendir ama bu çalışmada o kadar düşük olmadığından analizde tesadüfi hataların varlığını belirtir.

Yuvalanmış desen için kestirilen G ve Phi katsayı değerleri Tablo 9’da belirtilmiştir.

Tablo 9

<i>(bx(m:p)) Desenine Ait G ve Phi Katsayı Değerleri</i>	
Katsayı	Değer
G katsayısı	0,76
Phi katsayısı	0,64

Tablo 9’daki G ve Phi katsayıları incelendiğinde G katsayı değerinin 0,76, Phi katsayı değerinin ise 0,64 kestirildiği görülmektedir. G kuramında G ve Phi katsayılarının 0,80 ve 0,80’in üzerinde kestirildiğinde kabul edilebilir seviyelerdedir (Shavelson ve Webb, 1991). Bu sonuçlara göre elde edilen güvenilirlik katsayılarının istenilen düzeylere yakın olduğu görülmüştür. Sınavda sorulan soruların bireyler, maddeler ve puanlayıcılar arası değişkenlikleri yorumlamada yeterli görülebileceği söylenebilir.

(bx(m:p)) Deseninde Madde ve Puanlayıcı Sayılarının Artırılması ve Azaltılmasıyla Yapılan K Çalışması

40 bireyin olduğu bir testte farklı maddelerin farklı puanlayıcılar tarafından değerlendirmesinden oluşan veriye K çalışması uygulanmıştır. (bx(m:p)) deseninde bireyler ölçme objesi olarak belirlendiği, madde ve puanlayıcı sayılarının değiştirilmesiyle oluşturulan senaryolar için kestirilen G ve Phi katsayı değerleri, bağıl ve mutlak hata varyansları tablo 10’da yer almaktadır.

Tablo 10

(bx(m:p)) Deseni Madde ve Puanlayıcı Sayılarının Değiştirilmesiyle Kestirilen G ve Phi Katsayı Değerleri , Bağlı ve Mutlak Hata Varyansları

(bx(m:p))						
N_B	N_P	N_{MP}	G	Φ	σ²(δ)	σ²(Δ)
40	2	2	0,76	0,64	22,42	39,60
40	3	2	0,82	0,72	14,95	26,40
40	4	2	0,86	0,78	11,21	19,80
40	5	2	0,89	0,81	8,97	15,84
40	6	2	0,90	0,84	7,47	13,20
40	7	2	0,92	0,86	6,41	11,31
40	2	3	0,82	0,72	14,94	26,40
40	2	4	0,86	0,78	11,21	19,80
40	2	5	0,89	0,81	8,97	15,84
40	2	6	0,90	0,84	7,47	13,20
40	2	7	0,92	0,86	6,41	11,31
40	2	10	0,94	0,90	4,48	7,92
40	10	2	0,94	0,90	4,48	7,92
40	6	6	0,97	0,94	2,49	4,40
40	2	1	0,61	0,47	44,84	79,20
40	1	2	0,61	0,47	44,84	79,20

N_M : Madde Sayısı, N_P : Puanlayıcı Sayısı

Tablo 10 incelendiğinde, madde sayısı (N_M = 2), puanlayıcı sayısı (N_P = 2) iken 0,76 ve 0,64 olan G ve Phi katsayıları, madde sayısı (N_M = 2), puanlayıcı sayısı (N_P = 10) 0,94 ve 0,90 olmuştur ve .18 ve .26 artış göstermiştir. Dolayısıyla madde sayısı sabit tutulup (N_M = 2) puanlayıcı sayısı artırıldığında (2 ≤ N_P ≤ 10) G ve Phi katsayıları

artmaktadır. Tablo 10 incelendiğinde, madde sayısı ($N_M = 2$), puanlayıcı sayısı ($N_P = 2$) iken 0,76 ve 0,64 olan G ve Phi katsayıları, madde sayısı ($N_M = 10$), puanlayıcı sayısı ($N_P = 2$) olduğunda 0,94 ve 0,90 olmuştur ve .18 ve .26 artış göstermiştir. Dolayısıyla puanlayıcı sayısı ($N_P = 2$) sabit tutulup, madde sayısı artırıldığında ($3 \leq N_M \leq 10$) G ve Phi katsayıları artmaktadır. Puanlayıcı ve madde sayısının her ikisinin de artırıldığı ($N_M = 6, N_P = 6$) durumda ise G ve Phi katsayısı 0,97 ve 0,94 çıkmıştır. Madde ve puanlayıcı sayısı ($N_M = 2, N_P = 2$) katsayıya göre 0,21 ve 0,30 yükselmiştir. Yine Tablo 10 incelendiğinde, madde sayısı ($N_M = 2$) sabit tutulup puanlayıcı sayısı ($N_P = 1$) azaltıldığında G ve Phi katsayısının 0,61 ve 0,47 elde edilmiştir. Puanlayıcı sayısı ($N_P = 2$) sabit tutulup, madde sayısı ($N_M = 1$) azaltıldığında G ve Phi katsayısı 0,61 ve 0,47 elde edilmiştir. Bu değerlere göre madde ve puanlayıcı sayısının aynı sayıda azaltılmasıyla G ve Phi katsayıları .15 ve .17 düşüş göstermiştir. Aynı zamanda puanlayıcı sayısı ($N_P = 2$) sabit tutulup madde sayısı ($N_M = 5$) sadece bir artırıldığında G ve Phi katsayıları sırasıyla 0,89 ve 0,81 ile istenilen seviyelere gelmektedir.

Tablo 10'a göre madde sayısı sabit tutulup, puanlayıcı sayısının artırılması durumunda bağıl ($\sigma^2(\delta)$) ve mutlak ($\sigma^2(\Delta)$) hata varyansları incelendiğinde, puanlayıcı sayısındaki artışa bağlı olarak hata varyanslarının azaldığı görülmektedir. Madde sayısı ($N_M = 2$), puanlayıcı sayısı ($N_P = 10$) olduğu durumda bağıl ve mutlak hata varyansları ($\sigma^2(\delta) = 4,48$), ($\sigma^2(\Delta) = 7,92$), madde sayısı ($N_M = 10$), puanlayıcı sayısı ($N_P = 2$) olduğu durumda bağıl ve mutlak hata varyansları ($\sigma^2(\delta) = 4,48$), ($\sigma^2(\Delta) = 7,92$) çıkmıştır. Puanlayıcı ve madde sayısının her ikisinin de artırıldığı durumda ise bağıl ve mutlak hata varyansları da düşüş göstermektedir. Madde ve puanlayıcı sayısı ($N_M = 6, N_P = 6$) olduğunda bağıl ve mutlak hata varyansı ($\sigma^2(\delta) = 2,49$), ($\sigma^2(\Delta) = 4,40$), madde ve puanlayıcı sayısının ($N_M = 2, N_P = 2$) olduğu durumda elde edilen bağıl ve hata varyanslarına göre ($\sigma^2(\delta) = 22,42$), ($\sigma^2(\Delta) = 39,60$) bağıl hata varyansında 19,93, mutlak hata varyansında ise 35,20 değerinde bir düşüş olmuştur. Tablo 10'a göre puanlayıcı sayısı ($N_P = 2$) sabit tutulup madde sayısı ($N_M = 1$) azaltıldığında bağıl ve mutlak hata varyansları artış göstererek sırasıyla ($\sigma^2(\delta) = 44,84$), ($\sigma^2(\Delta) = 79,20$) çıkmıştır. Aynı şekilde madde sayısı ($N_M = 2$) sabit tutulup puanlayıcı sayısı ($N_P = 1$) azaltıldığında bağıl ve mutlak hata varyansları artış göstererek sırasıyla ($\sigma^2(\delta) = 44,84$), ($\sigma^2(\Delta) = 79,20$) çıkmıştır. Madde ve puanlayıcı sayısı ($N_M = 2, N_P = 2$) olduğu durumda elde edilen bağıl ve mutlak hata varyanslarına göre sırasıyla 22,42 ve 39,60 artış olmuştur.

Yukarıda yapılan analizlere göre testte kullanılan madde ve puanlayıcı sayısının ayrı ayrı artırılması güvenilirliği artırmaktadır. Ayrıca madde ve puanlayıcı sayısının aynı

anda artırılmasının güvenilirliği en fazla artıran senaryo olduğu görülmüştür. Bu sonuçlara göre puanlayıcı sayısının sabit tutulup madde sayısının artırılması güvenilirliği yükseltmede olumlu etki sağlamaktadır. İş gücü, zaman ve ekonomiklik bakımından puanlayıcı sayısının artırılmadığı, madde sayısının artırıldığı bir sınav senaryosu kurulmasının avantajlı olacağı düşünülmektedir.

(bxm_{xp}) ve (bx(m:p)) Desenleri ile Kestirilen G Çalışması Sonuçlarının Karşılaştırılması

Birey, madde ve puanlayıcılardan oluşan veriye göre elde edilen senaryolara göre (bxm_{xp}) ve (bx(m:p)) desenlerine ait varyans kaynağı, varyans ve toplam varyansı açıklama yüzdeleri Tablo 11’de gösterilmiştir.

Tablo 11

(bxm_{xp}) ve (bx(m:p)) Desenlerine Ait Varyans ve Toplam Varyans Açıklama Yüzdeleri

(bxm_{xp})			(bx(m:p))		
Varyans Kaynağı	σ^2	%	Varyans Kaynağı	σ^2	%
b	30,30	16,6	b	30,5	16,24
m	0,02	0,00	p	0,0	29,36
p	27,41	15,00	m:p	30,2	50,18
bm	1,20	0,70	bp	0,0	7,31
bp	116,94	64,10	b(m:p), e	39,4	14,18
mp	0,15	0,10			
bmp	6,42	3,50			
Toplam	184,44	100		100,1	100

b: Birey, m: Madde, p: Puanlayıcı

Tablo 11 incelendiğinde, toplam varyansı açıklama yüzdeleri birey değişkeni için, (bxm_{xp}) deseninde %16,6 ve (bx(m:p)) deseninde %16,24 kestirilmiştir. Bireylere ait varyans değerlerinin her iki desende de benzer çıkması bireylerden kaynaklanan değişkenliklerin eşit ölçüldüğünü göstermiştir. Sonuç olarak bireylerin her iki desende de yapılan analizde birbirinden farklılaştığı yorumu yapılabilir.

Tablo 11'deki deęerlere gre puanlayıcı ana etkisine ait varyans bileşenlerinin toplam varyansı açıklama yüzdeleri, (bxm xp) deseninde %15,00 ve (bx(m:p)) deseninde %29,36'dır. Puanlayıcıların her iki desende de farklılığa sebep olduğu görülmektedir ancak yuvalanmış desende puanlayıcı ana etkisinin varyans yüzdesi değeri daha fazla olduğu için puanlayıcılar arası tutarsızlık (bx(m:p)) deseninde daha fazla kestirilmiştir.

Aynı tabloya göre maddelere ait varyans bileşenlerinin toplam varyansı açıklama yüzdeleri, (bxm xp) deseninde %0,0 ve (bx(m:p)) deseninde %50,18 çıkmıştır. Çaprazlanmış desende maddelerin güçlük, kolaylık bakımından farklılaşmadığı yorumu yapılabilir. Yuvalanmış desende maddeler puanlayıcılar ile yuvalandığından maddelerin etkisi puanlayıcılardan bağımsız bir şekilde yorumlanamaz. Elde edilen varyans yüzdesi değerinin yuvalanmış desende büyük çıkmasının sebebi de puanlayıcıların, maddeleri puanlamalarında maddeden maddeye tutarsızlığa sebep olduğu yorumu yapılabilir.

(bxm xp) ve (bx(m:p)) desenlerinde kestirilen birey puanlayıcı ortak etkileşimine ait varyans açıklama yüzdeleri sırasıyla %64,10 ve %7,31'dir. Her iki desende de birey puanlayıcı ortak etkileşimlerinde farklılıkların olduğu puanlayıcıların bireyden bireye farklı puanlamada yaptıkları yorumu yapılabilir ancak sayısal deęerlere göre çaprazlanmış desende puanlayıcıların, puanlamalarında bireyden bireye daha yüksek oranda tutarlı olmadıkları sonucu çıkarılabilir.

Tablo 11'e göre artık varyansın açıklama yüzdeleri (bxm xp) deseninde %3,5, (bx(m:p)) deseninde ise %14,18 çıkmıştır. İlk desende birey puanlayıcı ortak etkileşiminin toplam varyansa etkisi çok yüksek olduğu için artık varyansta elde edilen bu değere katkısının birey ya da puanlayıcı deęişkenlerinden kaynaklandığı söylenebilir. (bx(m:p)) deseninde ise artık varyansın toplam varyansı açıklama yüzdesi madde puanlayıcı etkileşiminin etkisi, puanlayıcı ana etkisi ve birey ana etkisinden sonra dördüncü en büyük deęer alan varyanstır. Dolayısıyla bu analizde de artık varyansın birey ve puanlayıcıdan kaynaklanan rastgele bir deęişken olduğu yorumu yapılabilir.

(bxm xp) ve (bx(m:p)) Desenleri ile Madde ve Puanlayıcı Sayılarının Deęiştirilmesiyle Kestirilen Karar Çalışmaları Sonuçlarının Karşılaştırılması

Birey, madde ve puanlayıcılardan oluşan veriye göre elde edilen senaryolara göre (bxm xp) ve (bx(m:p)) desenleri ile madde ve puanlayıcı sayılarının artırılması ve azaltılmasıyla yapılan karar çalışmalarıyla kestirilmiş bağıl ve mutlak hata varyansları Tablo 12'de gösterilmiştir.

Tablo 12

(bxm_{xp}) ve (bx(m:p)) Desenleri için Madde ve Puanlayıcı Sayılarının Değiştirilmesiyle Oluşturulmuş Senaryolara Göre Elde Edilen Bağlı ve Mutlak Hata Varyansları

(bxm _{xp})			(bx(m:p))						
N _B	N _M	N _P	$\sigma^2(\delta)$	$\sigma^2(\Delta)$	N _B	N _P	N _{M:P}	$\sigma^2(\delta)$	$\sigma^2(\Delta)$
93	2	2	60,67	74,42	40	2	2	22,42	39,60
93	2	10	12,62	15,38	40	10	2	4,48	7,92
93	2	1	120,75	148,24	40	1	2	44,84	79,20
93	3	2	59,94	73,68	40	2	3	14,94	26,40
93	1	2	62,88	76,68	40	2	1	44,84	79,20
93	6	2	59,20	72,92	40	2	6	7,47	13,20

B: Birey, P: Puanlayıcı, M: Madde, $\sigma^2(\delta)$: Bağlı hata varyansı, $\sigma^2(\Delta)$: Mutlak Hata Varyansı

Tablo 12'ye göre (bxm_{xp}) ve (bx(m:p)) desenlerinin karar çalışmaları sonucunda kestirilen bağlı ($\sigma^2(\delta)$) ve mutlak ($\sigma^2(\Delta)$) hata varyansları karşılaştırılmıştır. (bxm_{xp}) deseni ile araştırmada kullanılan (N_B = 93, N_M = 2, N_P = 2) birey, madde ve puanlayıcı sayısı uygulanarak elde edilen bağlı ve mutlak hata varyansları $\sigma^2(\delta) = 60,67$, $\sigma^2(\Delta) = 74,42$ olarak hesaplanmıştır. Maddelerin puanlayıcılara yuvalandığı (bx(m:p)) deseninde (N_B = 40, N_M = 2, N_P = 2) kestirilen bağlı ve mutlak hata varyansları $\sigma^2(\delta) = 2,42$, $\sigma^2(\Delta) = 39,60$ olarak hesaplanmıştır. Bu durumda çaprazlanmış desen ile kestirilen bağlı ve mutlak hata varyansları, yuvalanmış desen ile kestirilen bağlı ve mutlak hata varyanslarına göre daha büyük çıkmıştır. Analizi yapılan iki desen için kestirilen bağlı ve mutlak hata varyanslarının farkı sırasıyla 38,25 ve 34,82 çıkmıştır. Bu değerlere göre yuvalanmış desende elde edilen analiz daha hatalardan arınık sonuçlar vermiştir.

Tablo 12'ye göre (bxm_{xp}) deseninde madde sayısı sabit iken puanlayıcı sayısı azaltıldığı durumda (N_B = 93, N_M = 2, N_P = 1) elde edilen bağlı ve mutlak hata varyansları sırasıyla $\sigma^2(\delta) = 120,75$ ve $\sigma^2(\Delta) = 148,24$, puanlayıcı sayısı artırıldığı durumda (N_B = 93, N_M = 2, N_P = 10) ise bağlı ve mutlak hata varyansları sırasıyla $\sigma^2(\delta) = 12,62$ ve $\sigma^2(\Delta) = 15,38$ olarak hesaplanmıştır. Maddelerin puanlayıcılara yuvalandığı (bx(m:p)) deseninde ise madde sayısı sabit iken puanlayıcı sayısı azaltıldığı durumda (N_B = 40, N_M = 2, N_P = 1) elde edilen bağlı ve mutlak hata varyansları sırasıyla $\sigma^2(\delta) = 44,84$ ve $\sigma^2(\Delta) = 79,20$, puanlayıcı sayısı artırıldığında (N_B = 40, N_M = 2, N_P = 10) ise bağlı ve mutlak

hata varyansları sırasıyla $\sigma^2(\delta) = 4,48$ ve $\sigma^2(\Delta) = 7,92$ olarak hesaplanmıştır. Bu sonuçlara göre madde sayısı sabit tutulup puanlayıcı sayısının azaltılması ile her iki desende de hata varyans değerleri artarken, puanlayıcı sayısı artırılması ile her iki desende de küçük varyans değerleri elde edilebilmektedir.

Tablo 12'ye göre (bxm xp) deseninde puanlayıcı sayısı sabit iken madde sayısı azaltıldığı durumda ($N_B = 93, N_M = 1, N_P = 2$) elde edilen bağıl ve mutlak hata varyansları sırasıyla $\sigma^2(\delta) = 62,88$ ve $\sigma^2(\Delta) = 76,68$, madde sayısı artırıldığında ($N_B = 93, N_M = 6, N_P = 2$) ise bağıl ve mutlak hata varyansları sırasıyla $\sigma^2(\delta) = 59,20$ ve $\sigma^2(\Delta) = 72,92$ olarak hesaplanmıştır. Maddelerin puanlayıcılara yuvalandığı (bx(m:p)) deseninde puanlayıcı sayısı sabit iken madde sayısı azaltıldığı durumda ($N_B = 40, N_M = 1, N_P = 2$) elde edilen bağıl ve mutlak hata varyansları sırasıyla $\sigma^2(\delta) = 44,84$ ve $\sigma^2(\Delta) = 79,20$, madde sayısı artırıldığında ($N_B = 40, N_M = 6, N_P = 2$) ise bağıl ve mutlak hata varyansları sırasıyla $\sigma^2(\delta) = 7,47$ ve $\sigma^2(\Delta) = 13,20$ olarak hesaplanmıştır. (bxm xp) deseninde puanlayıcı sayısının sabit iken madde sayısının artırılması ile bağıl ve mutlak hata varyanslarında kayda değer bir düşüş olmazken, (bx(m:p)) deseninde madde sayısının artırılması önemli düşüşler sağlamıştır.

Veriler bütünüyle incelendiğinde, farklı madde ve puanlayıcı sayısı senaryolarına göre (bx(m:p)) deseniyle elde edilen bağıl ve mutlak hata varyanslarının, (bxm xp) deseniyle elde edilen bağıl ve mutlak hata varyanslarına göre daha küçük kestirilme eğiliminde olduğu görülmektedir.

(bxm xp) ve (bx(m:p)) desenlerine uygulanan senaryolara göre yapılan K çalışmalarından elde edilen G ve Phi katsayı değerleri Tablo 13'de belirtilmiştir.

Tablo 13

(bxm xp) ve (bx(m:p)) Desenlerinin K Çalışması Sonuçlarına göre G ve Phi Katsayı Değerleri

(bxm xp)			(bx(m:p))						
N_B	N_M	N_P	G	Phi	N_B	N_{MP}	N_P	G	Phi
93	2	2	0,33	0,28	40	2	2	0,76	0,64
93	2	10	0,71	0,66	40	10	2	0,94	0,90
93	2	1	0,20	0,17	40	1	2	0,61	0,47

(Devam ediyor)

Tablo 13 (devam)

(bxm xp) ve *(bx(m:p))* Desenlerinin K Çalışması Sonuçlarına göre G ve Phi Katsayı Değerleri

(bxm xp)					(bx(m:p))				
N _B	N _M	N _P	G	Phi	N _B	N _{MP}	N _P	G	Phi
93	3	2	0,34	0,29	40	2	3	0,82	0,72
93	1	2	0,33	0,28	40	2	1	0,61	0,47
93	6	2	0,34	0,29	40	2	6	0,90	0,84

B: Birey, P: Puanlayıcı, M: Madde, G: G Katsayısı, Phi: Phi Katsayısı

(bxm xp) deseninde arařtırmada kullanılan veriler ($N_B = 93$, $N_M = 2$, $N_P = 2$) ile kestirilen G katsayısı 0,33 ve Phi katsayısı 0,28'dir. (bx(m:p)) deseninde arařtırmada kullanılan veriler ($N_B = 40$, $N_M = 2$, $N_P = 2$) G katsayısı 0,76 ve Phi katsayısı .64'tür. Bu deęerlere göre yuvalanmış (bx(m:p)) deseninde kestirilen G ve Phi katsayı deęerleri, çaprazlanmış (bxm xp) deseninde kestirilen G ve Phi katsayı deęerlerine göre daha büyüktür.

Tablo 13'e göre (bxm xp) deseninde madde sayısı sabit iken puanlayıcı sayısı azaltıldığı durumda ($N_B = 93$, $N_M = 2$, $N_P = 1$) kestirilen G ve Phi katsayıları sırasıyla 0,20 ve 0,17, puanlayıcı sayısı artırıldığında ($N_B = 93$, $N_M = 2$, $N_P = 10$) ise G ve Phi katsayıları sırasıyla 0,71 ve 0,66 olarak hesaplanmıştır. Maddelerin puanlayıcılara yuvalandığı (bx(m:p)) deseninde madde sayısı sabit iken puanlayıcı sayısı azaltıldığı durumda ($N_B = 40$, $N_M = 2$, $N_P = 1$) kestirilen G ve Phi katsayı deęerleriyle sırasıyla 0,61 ve 0,47, puanlayıcı sayısı artırıldığında ($N_B = 40$, $N_M = 2$, $N_P = 6$) ise G ve Phi katsayı deęerleri sırasıyla 0,90 ve 0,84 olarak hesaplanmıştır. Bu sonuçlara göre madde sayısı sabit iken puanlayıcı sayısının azaltılması ile her iki desende de küçük G ve Phi katsayıları kestirilirken puanlayıcı sayısının artırıldığı durumlarda ile her iki desende de G ve Phi katsayı deęerleri daha büyük kestirilebilmektedir.

Tablo 13 incelendiğinde, (bxm xp) deseninde puanlayıcı sayısı sabit iken madde sayısı azaltıldığı durumda ($N_B = 93$, $N_M = 1$, $N_P = 2$) G ve Phi katsayıları sırasıyla 0,33 ve 0,28, madde sayısı artırıldığında ($N_B = 93$, $N_M = 6$, $N_P = 1$) ise G ve Phi katsayıları sırasıyla 0,34 ve 0,29 olarak kestirilmiştir. Maddelerin puanlayıcılara yuvalandığı (bx(m:p)) deseninde puanlayıcı sayısı sabit iken madde sayısı azaltıldığı durumda ($N_B = 40$, $N_M = 1$, $N_P = 2$) kestirilen G ve Phi katsayıları sırasıyla 0,61 ve 0,47, madde sayısı

artırıldığında ($N_B = 40$, $N_M = 10$, $N_P = 2$) ise G ve Phi katsayıları sırasıyla 0,94 ve 0,90 olarak bulunmuştur. (bxm_{xp}) deseninde puanlayıcı sayısı sabit iken madde sayısının azaltılıp artırılması arařtırmada kullanılan ana deęerlere gre G ve Phi katsayılarını .01 ile ok az deęişkenlik gstermiştir. (bx(m:p)) deseninde ise puanlayıcı sayısı sabit iken madde sayısının azaltıldığı durmda G ve Phi katsayılarında sırasıyla 0,15 ve 0,17 deęerinde bir dşüşe sebep olmuştur. (bx(m:p)) deseninde ise puanlayıcı sayısı sabit iken madde sayısının artırıldığı durumda ise G ve Phi katsayılarında sırasıyla 0,28 ve 0,26 deęerinde bir artışa sebep olmuştur. (bxm_{xp}) deseninin de G ve Phi katsayıları puanlayıcı sayısı sabit tutulup madde sayısının artırılıp azaltılmasından byk oranda etkilenmezken, (bx(m:p)) desenin de G ve Phi katsayı deęerleri puanlayıcı sayısı sabit tutulup madde sayısının artırılıp azaltılmasından etkilenmektedir ve yuvalanmış desenin analizinde G ve Phi katsayıları daha byk kestirilmiştir.

Elde edilen veriler btnyle incelendiğinde, farklı madde ve puanlayıcı sayısı senaryolarına gre (bx(m:p)) deseniyle elde edilen G ve Phi katsayıları, (bxm_{xp}) deseniyle elde edilen G ve Phi katsayılarına gre byk ıkma eęilimindedir.

BÖLÜM 4

SONUÇ VE ÖNERİLER

Araştırmada personel alımı sınavının uygulandığı bireylerin verdiği açık uçlu cevapların iki puanlayıcı tarafından değerlendirilmesi ile elde edilen değerler G kuramına göre farklı desenler uygulanarak analiz edilmiştir. Bu analizlere göre uygulanan desenlerin G kuramı sonuçları karşılaştırılmış ve bu sonuçlara göre çok sayıda puanlayıcının kullanıldığı sınavlarda en az işgücü sarf ederek güvenilir sonuçlar elde etmek için puanlayıcıların ya da maddelerin etkisi belirlenmeye çalışılmıştır. Bu bölümde, çalışmada elde edilen sonuçlara göre G kuramında farklı desenlerin uygulanıp karşılaştırılmasına dayalı çalışmalar ve benzer çalışmalar için önerilere yer verilmiştir.

Sonuçlar

Araştırmada G kuramına göre uygulanan desenler ve bu desenlerin karşılaştırmasını içeren bulguların sonuçları bu kısımda verilmiştir. Araştırmanın bulgularında ulaşılan sonuçlar alt amaçların sırasına göre verilmiştir.

Birinci Alt Amaca İlişkin Sonuçlar

Personel alımı sınavında 93 bireyin iki maddeye verdiği açık uçlu cevapların iki puanlayıcı tarafından puanlanmasıyla elde edilen verilerin G çalışmasının sonuçlarına göre birey ve puanlayıcı ana etkisinin benzer varyans açıklama yüzdeleri olduğu ve analiz içinde yüksek varyans değerlerine sahip olduğu gözlemlenmiştir. Gugiu ve Gugiu (2012)'da çalışmalarında puanlayıcılardan kaynaklanan varyans değerinin yüksek olduğu sonucuna ulaşmaları bu çalışmanın sonuçlarını desteklediği görülmüştür. Madde ana etkisinin varyans açıklama yüzdesi sıfırdır. Alkan (2013) da çalışmasında madde ana etkisini sıfır elde etmiştir. Aynı zamanda Pekin (2015) de çalışmasında madde ana etkisini çok düşük bir değer elde etmiştir. Bu araştırmacıların yaptığı çalışmalar bu çalışmayı destekler niteliktedir. Birey-madde ortak etkisi varyans açıklama yüzdesi sıfıra yakındır. Birey-puanlayıcı ortak etkisi ise en büyük varyans açıklama yüzdesine sahiptir.

Dolayısıyla puanlayıcıların değerlendirmesi bireyden bireye değişkenlik göstermiştir. G ve Phi katsayıları ise sırasıyla 0,33 ve 0,29 olarak elde edilmiştir. Elde edilen değerlerin beklenen güvenilirlik seviyesinde olmadığı görülmektedir. Bu durum uygulanan testin bireyler ya da puanlayıcılar arası farklılığı iyi kestiremediği şeklinde yorumlanmıştır.

Testte kullanılan madde ve puanlayıcı sayılarının artırılıp azaltılarak istenilen güvenilirlik seviyesine gelip gelemeyeceğine yönelik çeşitli senaryolar oluşturulup karar (K) çalışmaları yapılmıştır. Karar çalışmasına göre tek başına madde sayısının artırılması güvenirliliğin artmasını sağlayamamaktadır. Puanlayıcı sayısının tek başına artırılması ise güvenirliliği artıran bir etki yapmıştır. Alharby (2006), çalışmasında puanlayıcı sayısının artırılmasının güvenirliliği artıran bir etki yaptığı sonucuna ulaşmıştır. Lane ve Sabers (1989)'da çalışmalarında puanlayıcı sayısını azaltmanın G ve Phi katsayısını düşüreceği aynı zamanda bağıl ve mutlak hata varyanslarını artıracığı sonucunu elde etmişlerdir. Deliceoğlu (2009), Büyükkıdık ve Anıl (2012), Alkan (2013) ve Pekin (2015)'de araştırmalarında uyguladıkları karar çalışmasının analizinde aynı bu çalışmada olduğu gibi puanlayıcı sayısı artırmanın madde sayısını artırmaya göre daha güvenilir sonuçlar elde edilebildiğine yönelik yorumlarda bulunmuşlardır. Puanlayıcı ve madde sayısının aynı anda artırılması da güvenirliliğe olumlu katkı sağlamaktadır. Alkahtani (2012), puanlayıcı ve madde sayısının artırılmasının güvenirliliğe olumlu katkı sağladığı sonucuna ulaşmıştır. Bu sonuçlara göre bu tip bir analizde iş gücü ve zaman faktörleri de göz önünde bulundurularak hem madde hem puanlayıcı sayısının güvenirliliğe katkı sağlaması amacıyla asgari düzeyde artırıldığı bir sınav düzenlenebileceği öngörülmüştür.

İkinci Alt Amaca İlişkin Sonuçlar

G çalışması sonucunda kestirilen varyans bileşenlerine göre birey ana etkisi ve artık varyans etkisinin benzer varyans açıklama yüzdeleri olduğu halde madde puanlayıcı etkisinin varyans açıklama yüzdesi en büyük varyans açıklama yüzdesine sahiptir. Puanlayıcı ana etkisinin varyans açıklama yüzdesi ise en büyük varyans açıklama yüzdesine sahip madde puanlayıcı etkisinden sonra ikinci en büyüktür. Bu durumda analizde puanlayıcılardan kaynaklanan bir değişkenlik etkisinin bulunduğu görülmüştür. Puanlayıcılar arasında maddeden maddeye farklılıklar oluşmuştur. Nalbantoğlu Yılmaz ve Gelbal (2011), çalışmalarında varyans açıklama yüzdelerinden görev ana etkisini fazla bulurken bu araştırmada puanlayıcı ana etkisi fazladır. Araştırmalarda aynı işlemler

yapılmış olsa da sonuçların farklı çıkmasının birey, madde ve puanlayıcı artık etkisinden kaynaklandığı düşünülmektedir.

G ve Phi katsayı değerleri sırasıyla 0,76 ve 0,64 elde edilmiştir. G ve Phi katsayılarının istenilen düzeylere yakın olarak elde edildiği görülmüştür. Yuvalanmış desen ile çaprazlanmış desene kıyasla daha büyük G ve Phi katsayı değerleri elde edildiği görülmektedir. Nalbantoğlu Yılmaz ve Gelbal (2011)'da araştırmalarında yuvalanmış desen için kestirilen G ve Phi katsayı değerlerinin daha yüksek çıkma eğiliminde olduğunu bulmuşlardır.

Analizde madde ve puanlayıcı sayısı artırılıp azaltılarak karar (K) çalışması yapılmıştır. Madde sayısı sabit tutulup puanlayıcı sayısının artırılması ile güvenilirlik artmaktadır. Puanlayıcı sayısı sabit tutulup madde sayısının artırılması ile de güvenilirlik artmaktadır. Benzer şekilde madde ve puanlayıcı sayısının aynı anda artırılması ile de güvenilirlik artırılabilir. Madde sayısının ($5 \leq N_M \leq 10$) aralıklarında, puanlayıcı sayısının ($5 \leq N_P \leq 10$) aralıklarında olduğu durumlarda G ve Phi katsayıları beklenen değerlerde olduğu görülmüştür. Aynı zamanda madde ve puanlayıcı sayısının ($N_M = 6, N_P = 6$) olduğu durumlarda G ve Phi katsayıları G kuramına göre beklenen değerlerde olduğu görülmüştür. Shavelson ve diğerleri (1990), çalışmalarında görev sayısını artırmanın daha güvenilir sonuçlar elde edilmesini sağlayacağını belirterek bu çalışmayı destekler nitelikte sonuçlara ulaşmışlardır. Elde edilen bu bilgilere göre puanlayıcı sayısının sabit tutulup madde sayısının artırılması güvenilirliği yükseltmektedir. Bu durumda iş gücü, zaman ve ekonomiklik bakımında puanlayıcı sayısının artırılmadığı madde sayısının artırıldığı bir senaryo ile testi uygulamanın avantajlı olacağı görülmüştür.

Üçüncü Alt Amaca İlişkin Sonuçlar

Çaprazlanmış ve yuvalanmış desenin G çalışması sonuçlarının karşılaştırılmasına göre birey ana etkisine yönelik varyans açıklama yüzdeleri çaprazlanmış ve yuvalanmış desenleri için birbirine yakın değerler almıştır. Her iki analizde de benzer sonuçlar elde edilmesi bireylerden kaynaklanan değişkenliklerin dolayısıyla bireyler arası farklılıkların benzer kestirilebildiğini göstermektedir. Puanlayıcı ana etkisinin her iki desende de farklılığa sebep olduğu görülmektedir ancak çaprazlanmış desende, yuvalanmış desene göre daha düşük varyans yüzdesi elde edilmiştir. Yuvalanmış desende daha fazla puanlayıcılar arası tutarsızlık elde edilmiştir. Puanlayıcı ana etkisinin her iki desende de

yüksek varyans etkisine sahip olduğunu Molloy ve Shimura (2005) ve Wang (2005)'da varyansın yüksek oranda bireysel farklılıklardan kaynaklandığını çalışmalarında desteklemişlerdir. Madde ana etkisi çaprazlanmış desende sıfır bulunarak maddeler zorluk düzeylerinin değişkenlik göstermediği sonucuna ulaşılmıştır. Yuvalanmış desende ise maddeler puanlayıcılar ile yuvalandığı için bağımsız yorumlanamamakla birlikte puanlayıcıların etkisi ile çok büyük varyans yüzdesine sahip olmuştur. Artık varyans değeri çaprazlanmış desende, yuvalanmış desene göre küçük çıkmıştır. Buna göre çaprazlanmış desenin yuvalanmış desene göre tesadüfi hatalardan daha arınık olduğu sonucu elde edilmiştir.

Çaprazlanmış ve yuvalanmış desenlere ait bağıl ve mutlak hata varyansları incelendiğinde yuvalanmış desenin daha küçük kestirildiği aynı şekilde çaprazlanmış ve yuvalanmış desenlere ait G ve Phi katsayı değerleri incelendiğinde yuvalanmış desene ait G ve Phi katsayı değerlerinin daha büyük kestirildiği görülmüştür. Bu farkın çaprazlanmış desende tek başına değerlendirilebilen madde ana etkisinin, yuvalanmış desende maddelerin değişkenlik etkisinin puanlayıcılar ile yuvalanmış şekilde değerlendirilmesinden kaynaklandığı düşünülmektedir. Yuvalanmış desende puanlayıcılardan kaynaklanan varyans farklı maddeleri değerlendirdikleri için G ve Phi katsayı değerlerinin, çaprazlanmış desene göre daha büyük kestirilebilmesine olanak sağladığı düşünülmektedir.

Dördüncü Alt Amaca İlişkin Sonuçlar

Madde sayısının sabit tutulup puanlayıcı sayısının artırıldığı senaryolarda her iki desende de bağıl ve mutlak hata varyanslarının azaldığı, puanlayıcı sayısının azaltıldığı senaryolarda ise bağıl ve mutlak hata varyanslarının arttığı gözlemlenmiştir. Puanlayıcı sayısının sabit tutulup madde sayısının artırıldığı ya da azaltıldığı senaryolarda çaprazlanmış desende bağıl ve mutlak hata varyanslarında önemli değişiklikler olmazken, yuvalanmış desende puanlayıcı sayısının sabit iken madde sayısının artırıldığı senaryo uygulandığında bağıl ve mutlak hata varyansları düşmekte, puanlayıcı sayısının sabit tutulup madde sayısının azaltıldığı senaryolarda ise bağıl ve mutlak hata varyansları artmaktadır. (bx(m:p)) deseninde hem puanlayıcı hem madde sayısının artırılması bağıl ve mutlak hata varyanslarının düşmesine sebep olmuştur. Lin ve Zhang (2014)'de çalışmalarında uygun G ve Phi katsayılarına ulaşabilmek için 3 ile 6 arasında puanlayıcı sayısı olmasını gerektiği yorumunu yapmışlardır. Genel olarak bağıl ve mutlak hata

varyans deęerlerinin yuvalanmış desende daha küçük kestirilme eğiliminde olduęu ve buna göre G ve Phi katsayılarının da daha yüksek katsayı deęerlerine ulaştığı gözlemlenmiştir. Bağcı (2015)'da araştırmasında bu araştırmada da uygulanan çaprazlanmış ve yuvalanmış deseni kendi çalışmasına uygulayarak benzer sonuçlar elde ederek yuvalanmış desende güvenilir sayılabilecek katsayılar elde edebilmek için uygulanacak senaryodaki yüzey sayılarının, daha ekonomik olduğunu bulmuştur. Bu araştırmada da yuvalanmış desenin uygulandığı durumda daha yüksek G ve Phi katsayıları elde edilmesi bu sonucu desteklemektedir.

Öneriler

Bu kısımda araştırma bulgularından elde edilen sonuçlara göre uygulayıcı ve araştırmacılara yönelik önerilere yer verilmiştir.

Uygulayıcılara Yönelik Öneriler

1. Puanlayıcılar arası tutarlılığın elde edilebilmesi testten daha güvenilir sonuçlar kestirilmesini sağlayacağı için puanlayıcı sayısının artırılması önerilmektedir. Puanlayıcı sayısının artırılmadığı durumlarda ise puanlayıcı güvenilirliğini sağlamak amacıyla puanlayıcıların geçerlilik ve güvenilirlik çalışmaları yapılmış bir dereceli puanlama anahtarı kullanmaları sağlanmalıdır. Ayrıca puanlama tutarlılığının sağlanabilmesi için puanlayıcılar arası güvenirliliğin test edileceği bir test düzeneği oluşturulmalıdır.
2. Teste girecek aday sayısının fazla olduğu durumlarda puanlayıcı sayısının artırılması mümkün olamamaktadır. Daha güvenilir sonuçlar elde edebilmek amacıyla testteki madde sayısının artırılması önerilmektedir. Madde sayısının artırılmasının mümkün olmadığı durumlarda ayıricılıkları daha yüksek kestirilen maddelerin bu tip sınavlarda kullanılması önerilmektedir.
3. Teste girecek aday sayısının fazla olduğu durumlarda puanlayıcı sayısını artırmadan maddelerin puanlayıcılar arasında dönüşümlü puanlanacağı bir test düzeneği oluşturmak testten güvenilir sonuçların elde edilmesi amacıyla önerilmektedir.

Arařtırmacılara Yönelik Öneriler

1. Çalışmada maddeleri puanlayıcılar ile yuvalandıđı bir senaryo ile G ve K çalışmaları yapılmıřtır. Bireylerin puanlayıcılar ile yuvalandıđı bir senaryoya G ve K çalışmasının uygulanması ileriki arařtırmalarda yer alabilir.
2. Çalışmada bireyler ölçme objesi olmak üzere madde ve puanlayıcı olarak iki yüzey belirlenmiřtir. Puanlama anahtarını da bir yüzey olarak belirleyip çok yüzeyli desenler uygulanabilir. Aynı zamanda sınava giren bireylerin lisans ya da yüksek lisans mezunu olma durumlarına göre deđiřkenlik durumları da analizde deđerlendirilebilir.



KAYNAKLAR

- Aiken, L. R. (2000). *Psychological testing and assessment*. (Tenth Edition). Boston: Allyn and Bacon.
- Alhaktani, S. F. (2012). *Oral Performance Scoring Using Generalizability Theory and Many-Facet Rasch Measurement: A Comparison Study*. (Unpublished doctoral dissertation) The Pennsylvania State University. Retrieved from https://etda.libraries.psu.edu/files/final_submissions/7551
- Alharby, E. (2006). *A Comparasion Between Two Scoring Methods, Holistic vs Analytic, Using Two Measurement Models, The Generalizability Theory and Many-Faced Rasch Measurement Within the Context of Performance Assesment*. (Unpublished doctoral dissertation). The Pennsylvania State University.
- Anastasi, A. (1982). *Psychological testing*. (Fifth Edition). New York: Macmillan Publishing.
- APA. (2014). *Standards for educational and psychological testing*. Washington: American Educational Research Association.
- Atılğan, H. (2004). *Genellenebilirlik Kuramı ve Çok Değişkenlik Kaynaklı Rasch Modelinin Karşılaştırılmasına İlişkin bir Araştırma* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi. Ankara, Türkiye.
- Atılğan, H., Kan, A. ve Doğan, N. (2007). *Eğitimde Ölçme ve Değerlendirme*. (2. Baskı). Ankara: Anı Yayıncılık
- Barnes, L. L. B. & Barnes, M. W. (1993). Academic discipline and generalizability of student evaluations of instruction. *Research in Higher Education*, 34, 135-49. <https://doi.org/10.1007/BF00992160>
- Başat, B. (2015). *6. Sınıf Fen Bilimleri Dersi Elektriği İletelim Ünitesinin Değerlendirilmesinde Kullanılan Ölçme Araçlarının Güvenirliğinin Genellenebilirlik Kuramına göre Analizi* (Yayımlanmamış Yüksek Lisans Tezi). Gaziosmanpaşa Üniversitesi. Tokat, Türkiye.
- Başol, G. ve Yüksel, M. (2017). Bir genellenebilirlik analizi çalışması: "Hücre Bölünmesi & Kalıtım Ünitesi". *International Journal of Assessment Tools in Education*. Vol 4(2), 149-165. <https://doi.org/10.21449/ijate.303991>
- Baykul, Y. (2010). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulama*. Ankara: Pegem Akademi.
- Baykul, Y ve Turgut, M. F. (2011). *Eğitimde Ölçme ve Değerlendirme*. (Üçüncü Basım). Ankara: Pegem Akademi.

- Bottema-Beutel, K., Lloyd, B., Carter, E. W. & Asmus, J. M. (2014). Generalizability and decision studies to inform observational and experimental research in classroom settings. *American Journal on Intellectual and Developmental Disabilities*, 119 (6),589–605. doi: 10.1352/1944-7558-119.6.589.
- Brennan, R. L. (2001). *Generalizability theory*. Iowa: Springer.
- Brown, D. J. (2005). Generalizability and decision studies. *JALT Testing & Evaluation SIG Newsletter*. 9(1): (p.12-16). Retrieved from <http://hosted.jalt.org/test/PDF/Brown21.pdf>
- Bruckner, C. T., Yoder, P. & McWilliam, R. (2006). Generalizability and decision studies: an example using conversational language samples. *Journal of Early Intervention* 28(2):139-153. DOI: 10.1177/105381510602800205
- Cardinet, J., Johnson, S. & Pini, G. (2010). *Applying generalizability theory using edug*. New York: Routledge Taylor & Francis Group.
- Cohen, R. J. & Swerdlik, M. E. (2005). *Psychological testing and assessment an introduction to tests and measurement*. (Sixth Edition). New York: The McGraw-Hill Companies.
- Crocker, L & Algina, J. (1986). *Introduction to classical and modern test theory*. Florida: Harcourt Brace Jovanovich College Publishers.
- Cronbach, L. J. (1990). *Essentials of psychological testing*. (Fifth Edition). New York: Harper&Row Publishers.
- Cronbach, L.J., Gleser, G.C., Nanda, H. & Rajaratnam, N. (1972). The dependability of behavioral measurements: Theory of generalizability of scores and profiles. New York: Wiley. doi: 10.1126/science.178.4067.1275
- Damlar Demirci, P. (2019). *Açık Uçlu Soruların Puanlama Yöntemlerinin Genellenabilirlik Kuramına göre İncelenmesi* (Yayımlanmamış Yüksek Lisans Tezi). Ege Üniversitesi. İzmir, Türkiye.
- Deliceoğlu, G. (2009). *Futbol Yetilerine İlişkin Dereceleme Ölçeğinin Genellenebilirlik ve Klasik Test Kuramına Dayalı Güvenirliliklerinin Karşılaştırılması* (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi. Ankara, Türkiye.
- Doğan, C. D. ve Anadolu, H. Ö. (2016) Genellenebilirlik Kuramında Tümüyle Çaprazlanmış ve Maddelerin Puanlayıcılara Yuvalandığı Desenlerin Karşılaştırılması. *Kastamonu Eğitim Dergisi*. Cilt: 25(1). 361-372. Erişim adresi <https://dergipark.org.tr/tr/download/article-file/297676>
- Eason, S. (1989). Why generalizability theory yields better results than classical test theory. Paper presented at the Annual Meeting of the MidSouth Educational Research Association (Little Rock, AR, November 8-10, 1989). (ERIC Document Reproduction Service No. ED 3 14 434). Retrieved from <https://eric.ed.gov/?id=ED314434>

- Erkuş, A. (2003). *Psikometri Üzerine Yazılar*. Ankara: Türk Psikologlar Derneği Yayınları.
- Erkuş, A. (2009). *Davranış Bilimleri için Bilimsel Araştırma Süreci*. (2. Baskı). Ankara: Seçkin Yayıncılık.
- Gao, X. & Brennan, R. L. (2001). Variability of estimated variance components and related statistics in a performance assessment. *Applied Measurement in Education* 14(2):191-203. doi: 10.1207/S15324818AME1402_5
- Gugiu, M, R. & Gugiu, C. (2012). Utilizing generalizability theory to investigate the reliability of the grades assigned to undergraduate research papers. *Journal of MultiDisciplinary Evaluation*, 8(19), 26-40. ISSN 1556-8180.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Güler (2008), *Klasik Test Kuramı, Genellenebilirlik Kuramı ve Rasch Modeli Üzerine Bir Araştırma* (Yayımlanmamış Doktora Tezi). Hacettepe Üniversitesi. Ankara, Türkiye.
- Güler (2011), Rastgele Veriler Üzerinde Genellenebilirlik Kuramı ve Klasik Test Kuramı'na Göre Güvenirliğin Karşılaştırılması. *Eğitim ve Bilim*. 36(162) 225-234. Erişim adresi <http://egitimvebilim.ted.org.tr/index.php/EB/article/view/993/323>
- Güler, N., Kaya Uyanık, G. ve Taşdelen Teker, G (2012). *Genellenebilirlik Kuramı*. Ankara. Pegem Akademi.
- Hagtvet, K. A., Frantisek, M. & Sharma, S. (2001). Generalizability of self-related cognitions in test anxiety. *Personality and Individual Differences*. 31(7), 1147-1171. [https://doi.org/10.1016/S0191-8869\(00\)00212-9](https://doi.org/10.1016/S0191-8869(00)00212-9)
- Hambleton, R. K. (2004). Theory, methods, and practices in testing for the 21st century. *Psicothema* 2004. 16(4), 696-701. Retrieved from www.psycothema.com
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. New York: Springer Netherlands.
- Han, C. (2018). A Generalizability theory study of optimal measurement design for a summative assessment of English/Chinese consecutive interpreting. *Language Testing*. 30(3), 419-438. doi: 10.1177/0265532218809396.
- Hintze, J. M & Matthews, W. J. (2004). The generalizability of systematic direct observations across time and setting: A preliminary investigation of the psychometrics of behavioral observation. *School Psychology Review* 33(2), 258-270. doi: 10.1080/02796015.2004.12086247.
- Hintze, J. M., Owen, S. V., Shapiro, E. S. & Daly III, E. J. (2012). Generalizability of oral reading fluency measures: Application of G theory to curriculum-based measurement. *School Psychology Quarterly* 15(1):52-68. doi: 10.1037/h0088778.

- Hopkins, K. D. (1998). *Educational and Psychological measurement and evaluation*. Boston: Allyn&Bacon.
- Karasar, N. (1998). *Araştırmalarda Rapor Hazırlama Yöntemi*. Ankara: Pars Matbaacılık.
- Kline, P. (2000). *The handbook of psychological testing*. London and New York: Routledge.
- Kubiszyn, T. & Borich, G. (2007). *Educational testing and measurement*. (8th Edition). Hoboken, NJ: John Wiley & Sons, Inc.
- Kunnan, J. A. (1992). An investigation of a criterion-referenced test using G-theory, and factor and cluster analyses. *Sage journals*. doi: 10.1177/026553229200900104
- Lakes, K. D. & Hoyt, W. T. (2009). Applications of generalizability theory to clinical child and adolescent psychology research. *Journal of Clinical Child Adolescent Psychology Research*. 38(1), 144-165. doi: 10.1080/15374410802575461
- Lane, S., & Sabers, D. (1989). Use of generalizability theory for estimating the dependability of a scoring system for sample essays. *Applied Measurement in Education*, 2(3), 195-205. doi: 10.1207/s15324818ame0203_1
- Lee, G. & Frisbie, D. A. (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets. *Applied Measurement in Education*, 12(3), 237-255. doi: 10.1207/S15324818AME1203_2
- Lee, Y. & Kantor, R. (2007). Evaluating prototype tasks and alternative rating schemes for a new ESL writing test through G theory. *International Journal of Testing*, 7(4), 353-385. doi: 10.1080/15305050701632247.
- Li, M. Shavelson, R. J., Yin, Y & Wiley, E. (2015). Generalizability theory. *The Encyclopedia of Clinical Psychology*, 1-19 doi: 10.1002/9781118625392.wbecp352.
- Lin, C. & Zhang, J. (2014). Investigating correspondence between language proficiency standards and academic content standards: A generalizability theory study. October 2014. *Language Testing* 31(4), 413-431. doi: 10.1177/0265532213520304.
- Molloy, H. & Shimura, M. (2005). An examination of situational sensitivity in medium-scale interlanguage pragmatics research. In T Newfields, Y. Ishida, M. Chapman, & M. Fujioka (Eds.) Proceedings of the May. 22- 23, 2004 *JALT Pan-SIG Conference Tokyo: JALT Pan SIG Committee*.(p.16 -32). Retrieved from <http://hosted.jalt.org/pansig/2004/HTML/ShimMoll.htm>
- Mushquash, C. & O'Connor, B. P. (2006). SPSS and SAS programs for generalizability theory analyses. *Behavior Research Methods*. 38 (3), 542-547. doi: 10.3758/BF03192810

- Nalbantoğlu, F. (2009). *Performans Ölçümlerinde Genellenebilirlik Kuramıyla Farklı Desenlerin Karşılaştırılması* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi. Ankara, Türkiye.
- Nalbantoğlu Yılmaz, F. (2012). *Genellenebilirlik Kuramında Dengelenmiş ve Dengelenmemiş Desenlerin Karşılaştırılması-İntramuskuler Enjeksiyon Yapma İstasyonu Verileri Üzerinde Bir Uygulama* (Yayınlanmamış Doktora Tezi). Ankara Üniversitesi. Ankara, Türkiye.
- Nalbantoğlu Yılmaz, F ve Gelbal, S. (2011). İletişim becerileri istasyonu örneğinde genellenebilirlik kuramıyla farklı desenlerin karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi (H.U Journal of Education)*.41 (2011), 509-518. Erişim adresi <http://www.efdergi.hacettepe.edu.tr/yonetim/icerik/makaleler/718-published.pdf>
- Ohta, R, Plakans, L. M, & Gebril, A. (2018), Integrated writing scores based on holistic and multi-trait scales: A generalizability analysis. *Elsevier Assessing Writing* 38 (2018), 21-36.
- Pekin, Z. (2015). *Otizm Sosyal Beceriler Profili Ölçeğinde Puanlayıcılar Arası Güvenirliğin Klasik Test Kuramı ve Genellenebilirlik Kuramına Göre Karşılaştırılması* (Yayımlanmamış Yüksek Lisans Tezi). Hacettepe Üniversitesi. Ankara, Türkiye.
- Pekin, Z., Çetin, S., & Güler, N. (2015). Comparison of interrater reliability based on different theories for autism social skills profile. *Journal Of Measurement And Evaluation In Education And Psychology-Epod. Vol.9*, pp.202-215. 2018. doi: 10.21031/epod.388590
- Popham, J. W. (1990). *Modern educational measurement: A practitioner's perspective*. New Jersey: Prentice Hall.
- Shavelson, R. J., Li, W., Mayberry, P. W. & Webb, N. (1990). Generalizability of job performance measurements: Marine corps rifleman. *Military Psychology*, 2(3), 129-144.
- Shavelson, R. J. & Webb, N. M. (1991). *Generalizability theory a primer*. California: Sage Publications.
- Shavelson, R. J. & Webb, N. M. (2005). *Generalizability theory*. Retrieved from https://web.stanford.edu/dept/SUSE/SEAL/Reports_Papers/methods_papers/G%20Theory%20AERA.pdf
- Shavelson, R.J., Webb, N. & Rowley, G. (1988). Using generalizability theory in counseling and development. July 1988. *Measurement and Evaluation in Counseling and Development*, 21(2), 81-90. Retrieved from https://www.researchgate.net/publication/232540065_Using_Generalizability_Theory_in_Counseling_and_Development

- Suen, H. K. (1990). *Principles of test theories*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Sun, A. Valiga, M. J. & Gao, X. (1997). Using generalizability theory to assess the reliability of student ratings of academic advising. *The Journal of Experimental Education*, 65(4), 367-379, summer 1997. doi: 10.1080/00220973.1997.10806611
- Thorndike, R. M. (2005). *Measurement and evaluation in psychology and education*. (Seventh Edition). New Jersey: Pearson Education, Inc.
- Turgut, M. F. (1990). *Eğitimde Ölçme ve Değerlendirme Metotları*. Ankara: Nüve Matbaası.
- VanLeeuwen, D. M. (1997). Assessing reliability of measurements with generalizability theory: An application to inter-rater reliability. *Journal of Agricultural Education*. 38(3), 36-42. ISSN: ISSN-1042-0541
- Yelboğa, A. ve Tavşancıl, E. (2010). Klasik Test ve Genellenebilirlik Kuramına göre Güvenirliğin bir İş Performansı Ölçeği Üzerinde İncelenmesi. *Educational Sciences: Theory & Practice*, 10(3), 1824- 1854. ISSN: 1303-0485 / 2148-7561
- Wang, Z. (2005). *Estimating reliability a generalizability theory model for writing scores in C-base*. (Unpublished master's thesis). University of Missouri. Columbia, USA.
- Wasserman, R. H., Levy, K. N. & Loken, E. (2009). Generalizability theory in psychotherapy research: The impact of mutiple sources of variance on the dependibility of psychotherapy process ratings. *Psychotherapy Research*, 19(4-5), 397-408. doi: 10.1080/10503300802579156

EK-1: ASYM'DEN ALINAN VERİ KULLANIMI İÇİN İZİN BELGESİ



T.C.
ANKARA ÜNİVERSİTESİ REKTÖRLÜĞÜ
Sınav Yönetim Merkezi Birimi

Sayı : 59396580-622.03-E.3011

15.01.2019

Konu : Veri Talebi Hk.

SINAV YÖNETİM MERKEZİ BİRİMİNE

ASYM'de hazır bulunan verileri, adayların kişisel hiçbir bilgisi verilmeden Üniversitemiz gözetim ve denetiminde sadece bilimsel amaçlı kullanmak üzere talep eden araştırmacıların dilekçesi incelenmiş olup, söz konusu verilerin araştırmacılarla paylaşılması talebi Koordinatörlüğünüzce uygun bulunmuştur.

Bilgilerinizi saygı ile rica ederim.

e-İmza

Prof. Dr. Mualla SELÇUK
Kurumlar Arası Sınav Genel Koordinatörü

Ek : Dilekçe (2 sayfa)

Siz: 3079 sayılı Elektronik İmza Kanunu gereği bu belge elektronik imza ile onaylanmıştır.

Kırkkonuk Mahallesi Lale Sokak No:5 06430 Maltepe ANKARA TÜRKİYE

Telefon No: 0 312 222 05 39 - 0 312 222 05 40 Faks: 0312 222 05 31

E-posta: sayim@ankara.edu.tr Web adresi: <http://asyim.ankara.edu.tr> Twitter: @AnkaraUni

Bilgi için Şefik Müst. GÜMÜŞOY

Bilgi İht.

Telefon No: 0312 222 05 39-2900

BENZERLİK BİLDİRİMİ

“Personel Alımında Kullanılan Bir Yazılı Sınav Sonucunun Genellenebilirlik Kuramındaki Farklı Desenlerle Karşılaştırılması” başlıklı tezimin ana bölümü (ön bölüm, kaynaklar ve ekler hariç) Turnitin İntihali Engelleme Programı aracılığıyla incelenmiş ve ilgili rapor danışmanım tarafından da kontrol edilmiştir. Kontrol sırasında (1) “Beş sözcükten daha az olan benzeşmeler” (2) “Kaynaklar” (3) “Doğrudan Alıntılar” dışarıda tutulmuştur. Benzerlik kontrolüne ilişkin rapordan elde edilen sonuçlar aşağıda sunulmuştur.

Rapor Tarihi	01.12.2020
Gönderim Numarası	1461440023
Sayfa Sayısı	84
Sözcük Sayısı	16162
Karakter Sayısı	125,568
Benzerlik Oranı	%4
Savunma Tarihi	18.12.2020

Yukarıda belirtilen sonuçları gösteren Turnitin İntihali Engelleme Programı’na ilişkin orijinal raporu, sonuçlarda herhangi bir değişiklik yapmaksızın bu beyanım ekinde Enstitüye teslim ettiğimi, tezimin %10’dan fazla benzerlik oranı içerdiğinin belirlenmesi durumunda, bundan doğabilecek tüm yasal sorumluluğu kabul ettiğimi bildirir, saygılarımı sunarım.

Öğrencinin Adı Soyadı: İrem DEMİRBILEK ZORBA

Tarih: 01.12.2020

İmza: 

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı : İrem DEMİRBILEK ZORBA
E-Posta Adresi : iremdemirbilek@hotmail.com

İş Deneyimi :

Unvan	Görev Yeri	Yıl
Fen Bilgisi Öğretmeni	Milli Eğitim Bakanlığı (Hakkari, Edirne ve Ankara Milli Eğitim Müdürlükleri)	Ocak 2008-

Akademik Bilgiler

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	Fen Bilgisi Öğretmenliği	Eskişehir Osmangazi Üniversitesi	2007